Singapore Management University
Institutional Knowledge at Singapore Management University

Research Collection School Of Economics | School of Economics

9-2006

Set Inference for Semiparametric Discrete Games

Kyoo-il KIM
Singapore Management University, kyookim@smu.edu.sg

SMU ECONOMICS & STATISTICS
WORKING PAPER SERIES

Set Inference for Semiparametric Discrete Games

Kyoo il Kim

September 2006

Paper No. 16-2006

Set Inference for Semiparametric Discrete Games

Kyoo il Kim*

School of Economics and Social Sciences
Singapore Management University

September 26, 2006

Abstract

We consider estimation and inference of parameters in discrete games allowing for multiple equilibria, without using an equilibrium selection rule. We do a set inference while a game model can contain infinite dimensional parameters. Examples can include signaling games with discrete types where the type distribution is nonparametrically specified and entry-exit games with partially linear payoffs functions. A consistent set estimator and a confidence interval of a function of parameters are provided in this paper. We note that achieving a consistent point estimation often requires an information reduction. Due to this less use of information, we may end up a point estimator with a larger variance and have a wider confidence interval than those of the set estimator using the full information in the model. This finding justifies the use of the set inference even though we can achieve a consistent point estimation. It is an interesting future research to compare these two alternatives: CI from the point estimation with the usage of less information vs. CI from the set estimation with the usage of the full information.

Keywords: Semiparametric Estimation, Set Inference, Infinite Dimensional Parameters, Inequality Moment Conditions, Signaling Game with Discrete Types

JEL Classification: C13, C14, C35, C62, C73

1 Introduction

The econometric modeling of game theories has been of significant interest over the last decade including simultaneous games with complete information (Bjorn and Vuong (1984, 1985), Bresnahan and Reiss (1990, 1991), Tamer (2003), Bajari, Hong and Ryan (2004)) or with incomplete information (Brock and Durlauf (2001, 2003), Seim (2002), Sweeting (2004), Aradillas-Lopez (2005)), dynamic games (Aguirregabiria and Mira (2003), Bajari, Benkard, and Levin (2003), Berry, Ovstrovsky, and Pakes (2003), Pesendorfer and Schmidt-Dengler (2003)), and signaling games (Kim (2006)). Here we focus on static discrete games. For these games, depending on the equilibrium properties, a researcher can face with the issue of multiple equilibria. Several resolutions have been proposed such as imposing equilibrium selection rules[1] and redefining the space of outcomes in a game[2].

*Correspondence: Kyoo il Kim, School of Economics and Social Sciences, Singapore Management University, 90 Stamford Road, Singapore 178903, Tel: +65 6828–0876, Fax: +65 6828-0833, E-mail: kyookim@smu.edu.sg.

[1] Examples include Bjorn and Vuong (1984, 1985), Kooreman (1994), and Bajari, Hong and Ryan (2004) for games with complete information, Sweeting (2004) for a game with incomplete information, and Kim (2006) for signaling games with discrete types.

[2] See Bresnahan and Reiss (1990, 1991).

Alternatively, inspired by important work by Manski and co-authors (Manski (1990), Horowitz and Manski (1995), and Manski and Tamer (2002)) on bound analysis, some researchers have started to develop set inferences rather than a point estimation, without attempting to resolve the equilibrium selection (Sutton (2000), Ciliberto and Tamer (2003), and Andrews, Berry, Jia (2004) [ABJ]). In particular, we consider the model where some asymptotic inequalities may define a region of parameters rather than a single point in the parameter space. By definition, when there are multiple equilibria, there exist regions of unobservables that are consistent with the necessary conditions for more than one equilibrium. Therefore, the probability implied by the necessary condition for a given event is greater than or equal to the true probability of the event and a set inference including this paper utilizes these inequality conditions.

Another thing we note in the literature of the set inference is that parameters allowed in game models are only finite dimensional even though infinite dimensional parameter is naturally included in the model (For example, see Kim (2006)) or misspecification of a fully parametric model is concerned. This paper considers a set inference with infinite dimensional parameters. A consistent set estimator and a confidence interval are provided in the paper.

Our proposed set estimation and inference requires a consistent profiled estimator for the infinite dimensional parameters. An interesting case we note in this paper is that sometimes we can achieve a consistent point estimation of all the parameters including finite and infinite dimensional ones by losing some information in the model. For example, in Bresnahan and Reiss (1990, 1991), we disregard the information about which firms enter the market since we redefine the outcome space in terms of how many firms in the market. Due to this omitted information, we may end up point estimators whose variances are larger and thus have wider confidence intervals than those of the set estimator using the full information in the model. Comparison of these two will be also interesting.

The organization of this paper is as follows. Section 2 introduces the model we study. In Section 3, we extend the set estimator of ABJ to the semiparametric case and provide two examples of such models. In Section 4, we show the consistency of the set estimator. In Section 5, we propose a set inference. We conclude in Section 6. Technical details and mathematical proofs are presented in Appendix.

2 Model

Let Y_p be player p's action (or strategy) and X_p be a vector of player p's characteristics for $p = 1, \ldots, \mathfrak{p}$. Let ε_p be player p's unobservable to econometricians, which is a part of player p's payoff functions. We let $Y = (Y_1', Y_2', \ldots, Y_{\mathfrak{p}}')' \in \mathbb{R}^{l_1+\ldots+l_{\mathfrak{p}}}$, $X = (X_1', X_2', \ldots, X_{\mathfrak{p}}')' \in \mathbb{R}^k$ and let $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_{\mathfrak{p}})' \in \mathbb{R}^{\mathfrak{p}}$. Also let $\alpha_0 \equiv (\theta_0, h_0)$ be the parameters of interest in a game. Then, payoffs of the game are given as functions of $(Y, X, \varepsilon, \alpha_0)$ and an equilibrium of the game can be characterized comparing those payoffs for different actions or strategies. α_0 consists of a vector of finite dimensional unknown parameters, denoted by θ_0 and a vector of infinite dimensional unknown functions, denoted by h_0. Here we let h_0 be functions of X or subset of X alone without loss of generality since Y is discrete. We allow h_0 can depend on the parameters θ. We also let $\mathbb{S}(W)$ denote the support of the distribution of random variable W.

Now denote $\Omega(Y = y, X = x, \alpha)$ to be the region of ε under which Y takes the value y given $X = x$ and α. To be precise, $\Omega(y, x, \alpha) \equiv \Omega(Y = y, X = x, \alpha) = \{\varepsilon | Y_1 = y_1, Y_2 = y_2, \ldots, Y_{\mathfrak{p}} = y_{\mathfrak{p}}$ given $X = x$ and $\alpha\}$. Then, the probability that the necessary conditions for $Y = y$ holds, denoted by $P(y|x, \alpha)$, will equal the probability that ε belongs to $\Omega(y, x, \alpha)$ given $X = x$ and α. Depending on the game of interest and the equilibrium property of the game, a researcher can construct $\Omega(y, x, \alpha)$ accordingly. Examples can be found in ABJ for entry games and Kim (2006) for signaling games. Note that

$$P(y|x, \alpha) = \Pr(\varepsilon \in \Omega(y, x, \alpha)) \tag{1}$$

and thus the analytic form of $P(y|x, \alpha)$ will be given as long as the distribution of ε is assumed to be known. When $\alpha = \alpha_0$, this is a simple ε-orthant probability. Note that at the true parameter value $\alpha = \alpha_0$, the probabilities of the necessary conditions must be at least as large as the true probabilities of the events $y \in \mathbb{S}(Y)$ given $X = x$, denoted by $P_0(y|x)$:

$$P(y|x, \alpha_0) \geq P_0(y|x),\ \forall (y, x) \in \mathbb{S}(Y) \times \mathbb{S}(X). \tag{2}$$

Notice that this inequality follows from the fact that the outcome y implies the necessary conditions for y but the necessary condition need not imply the outcome y. Note that the inequalities in (2) are satisfied for the true α_0 and possibly for other values. It is possible that only one α satisfies the inequalities and that if the necessary conditions are derived from an incorrect model, then perhaps no α will satisfy the inequalities.

Now let $\mathcal{A}_0 \equiv \Theta_0 \times \mathcal{H}_0$ denote the asymptotically identified set of parameters that satisfy the inequality restrictions in (2). This $\mathcal{A}_0$ is the object we are trying to estimate from the model[3]. We may estimate both the finite dimensional parameters and the infinite dimensional parameters simultaneously. Alternatively, we may obtain consistent profiled estimates of $h_0(\cdot, \theta)$ from the model or an auxiliary model and then estimate θ_0 (thus $\alpha_0 = (\theta_0, h_0(\cdot, \theta_0))$ in the main estimation. Here we adopt the second approach where consistent estimates of $h(\cdot, \theta)$ for all $\theta \in \Theta$ are available. We will suppress the arguments of h_0 for notational convenience such that $(\theta, h) \equiv (\theta, h(\cdot, \theta))$, $(\theta, h_0) \equiv (\theta, h_0(\cdot, \theta))$, and $(\theta_0, h_0) \equiv (\theta_0, h_0(\cdot, \theta_0))$.

3 Set Estimator

Here we take the approach by ABJ. Noting ABJ only allows for finite dimensional parameters by construction[4], we adopt the second step estimation for infinite dimensional parameters where profiled estimates for infinite dimensional parameters are available in the pre-step and thus in the main estimation, we only deal with finite dimensional parameters. To focus on the treatment of the infinite dimensional parameters in this paper, we simplify discussions regarding the construction of estimators and related issues. Such issues can be found in ABJ. Here we assume that the model probabilities $\{P(y|X_i, \alpha\} : i = 1, \ldots, n\}$ have analytic closed form solutions.[5] Now we briefly review the data-dependent construction of X cells following ABJ[6]. Noting the data-dependent selection of X cells[7] will affect the asymptotic distribution of the statistics, we account for this dependency in the determination of the critical values later. Consider a set $\{q_\gamma : \gamma \in \Gamma\}$ of real-valued weight functions on $\mathbb{S}(X)$, where γ is a subset of $\mathbb{S}(X)$ and Γ is a collection of subsets of $\mathbb{S}(X)$. In particular, for each $y^{(j)} \in \mathbb{S}(Y) = \{y^{(1)}, \ldots, y^{(J)}\}$, we consider such $\mathcal{M}_j$ subsets of $\mathbb{S}(X)$ indexed by $\gamma_{j,m}$, $m = 1, \ldots, \mathcal{M}_j$. We let $\Gamma = \left\{\gamma_{j,m} \subset \mathbb{S}(X) : (j,m) \epsilon \mathcal{I}_{J,\mathcal{M}}\right\}$, where $\mathcal{I}_{J,\mathcal{M}} = \{(j,m) : m = 1, \ldots, \mathcal{M}_j, j = 1, \ldots, J \text{ with } J = l_1 \times \ldots \times l_{\mathfrak{p}}\}$. The functions $\{q_\gamma : \gamma \in \Gamma\}$ aggregate and/or weight the necessary condition for an equilibrium over different values of x. Now let $\widehat{\Gamma}_n = \left\{\widehat{\gamma}_{n,j,m} \subset \mathbb{S}(X) : (j,m) \in \mathcal{I}_{J,\mathcal{M}}\right\}$, where $\widehat{\gamma}_{n,j,m}$ is a random subset of $\mathbb{S}(X)$. For the consistency of the set estimator, we require $\widehat{\Gamma}_n \underset{p}{\rightarrow} \Gamma_0$ under certain metric described later where $\Gamma_0 = \left\{\gamma_{0,j,m} \subset \mathbb{S}(X) : (j,m) \in \mathcal{I}_{J,\mathcal{M}}\right\}$. Now we extend ABJ to the semiparametric case where a profiled consistent estimator of $h_0(\cdot, \theta)$, denoted by

[3] Note that $\mathcal{A}_0$ could be (i) the null set, (ii) a single point, (iii) a strict subset of the parameter space consisting of more than one point, or (iv) the entire parameter space. ABJ refers that the model is (i) rejected, (ii) point identified, (iii) set identified, or (iv) completely uninformative.

[4] It is because ABJ utilizes finite numbers of cells to facilitate the estimation, which is not compatible with infinite dimensional parameters. Simply it violates the order condition for identification.

[5] The model probabilities induced by the games may not have analytic closed form expressions. In that case we need to consider the simulated version of the probabilities which are not hard to construct in many cases. The analysis here can easily adopt the simulated version of model probabilities.

[6] We may also need to consider such cells for Y when the dimension of Y is high but here we implicitly assume that we do not have such a problem.

[7] Examples of constructing these cells and some efficiency issue can be found in ABJ.

$\widehat{h}(\cdot,\theta)$, is available for all $\theta \in \Theta$. Define

$$\begin{aligned} c_0(j,\gamma,\theta,h) &= \int \left(P(y^{(j)}|x,\theta,h(\cdot,\theta)) - P_0(y^{(j)}|x) \right) q_\gamma(x) dF_X(x) \text{ and} \\ \widehat{c}_n(j,\gamma,\theta,h) &= n^{-1}\sum_{i=1}^{n} \left(P(y^{(j)}|X_i,\theta,h(\cdot,\theta)) - \mathbf{1}[Y_i = y^{(j)}] \right) q_\gamma(X_i). \end{aligned} \tag{3}$$

Note that $E\left[\widehat{c}_n(j,\gamma,\theta,h)\right] = c_0(j,\gamma,\theta,h)$ for all (j,γ,θ,h) by construction. Hence, with iid observations, we have $\widehat{c}_n(j,\gamma,\theta,h) \underset{p}{\to} c_0(j,\gamma,\theta,h)$ provided that $c_0(j,\gamma,\theta,h)$ is well-defined. Necessary conditions for θ to be the true parameters are

$$P(y|x,\theta,h_0(\cdot,\theta)) - P_0(y|x) \geq 0, \ \forall (y,x) \in \mathbb{S}(Y) \times \mathbb{S}(X) \tag{4}$$

which implies that

$$c_0(j,\gamma_{0,k,m},\theta,h_0(\cdot,\theta)) \geq 0, \ \forall (j,m) \in \mathcal{I}_{J,\mathcal{M}}. \tag{5}$$

Define

$$\Theta_0 = \{\theta \in \Theta : (4) \text{ holds}\} \text{ and } \Theta_+ = \{\theta \in \Theta : (5) \text{ holds}\}.$$

By definition, the set Θ_0 is the smallest set of parameter values that necessarily includes the true value θ_0 (and thus $\alpha_0 \in \Theta_0 \times \mathcal{H}_0$). By construction, $\Theta_+ \supset \Theta_0$ since (4) implies (5). Now suppose that we have an initial nonparametric estimator $\widehat{h}(\cdot,\theta)$ of $h_0(\cdot,\theta)$ for each θ. Then we define a set estimator $\widehat{\Theta}_n$ of Θ_+ in the spirit of ABJ. To do that, we first define the estimator criterion function

$$Q_n(\theta,h) = \sum_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} \left|\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\theta,h)\right| \cdot \mathbf{1}[\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\theta,h) \leq 0] \tag{6}$$

whose population version of the criterion function is given by

$$Q(\theta,h) = \sum_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} \left|c_0(j,\gamma_{0,j,m},\theta,h)\right| \cdot \mathbf{1}[c_0(j,\gamma_{0,j,m},\theta,h) \leq 0]. \tag{7}$$

Note that the function $Q(\theta,h)$ is minimized and equals zero for all values (θ,h) which satisfy the necessary conditions $c_0(j,\gamma_{0,j,m},\theta,h) \geq 0$ for all $(j,m) \in \mathcal{I}_{J,\mathcal{M}}$, which implies that

$$\Theta_+ = \left\{\theta \in \Theta : \theta \text{ minimizes } Q(\theta,h_0(\cdot,\theta)) \text{ over } \Theta\right\}. \tag{8}$$

This justifies the construction of the set estimator $\widehat{\Theta}_n$ to be

$$\widehat{\Theta}_n = \left\{\theta \in \Theta : \theta \text{ minimizes } Q_n(\theta,\widehat{h}(\cdot,\theta)) \text{ over } \Theta\right\}, \tag{9}$$

where $\widehat{h}(\cdot,\theta) \in \mathcal{H}_n$ and $\mathcal{H}_n$ is a space of sieves that approximates $\mathcal{H}$ satisfying $\mathcal{H}_n \subseteq \mathcal{H}_{n+1} \subseteq \mathcal{H}$ for all $n \geq 1$. Note that if there exists a value of $(\theta,\widehat{h}(\cdot,\theta))$ for which $\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\theta,\widehat{h}) \geq 0$ for all $(j,m) \in \mathcal{I}_{J,\mathcal{M}}$, then $\widehat{\Theta}_n$ equals to

$$\left\{\theta \in \Theta : \widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\theta,\widehat{h}(\cdot,\theta)) \geq 0, \ \forall (j,m) \in \mathcal{I}_{J,\mathcal{M}}\right\}. \tag{10}$$

It is possible that the set defined in (10) is empty by the randomness in the estimator of $\widehat{c}_n(\cdot)$. We will let $\mathcal{A}_0 \equiv \Theta_0 \times \mathcal{H}_0$, $\mathcal{A}_+ \equiv \Theta_+ \times \mathcal{H}_+$, and $\widehat{\mathcal{A}}_n \equiv \widehat{\Theta}_n \times \widehat{\mathcal{H}}_n$ where $\mathcal{H}_0 = \{h \in \mathcal{H} : h = h_0(\cdot,\theta) \text{ and } \theta \in \Theta_0\}$, $\mathcal{H}_+ = \{h \in \mathcal{H} : h = h_0(\cdot,\theta) \text{ and } \theta \in \Theta_+\}$ and $\widehat{\mathcal{H}}_n = \{h \in \mathcal{H}_n : h = \widehat{h}(\cdot,\theta) \text{ and } \theta \in \widehat{\Theta}_n\}$, respectively.

The existence of $\widehat{\Theta}_n$ (i.e. the existence of $\widehat{\mathcal{A}}_n$) is guaranteed since $\widehat{\Theta}_n$ is defined as the collection of arguments that minimize a continuous function on a compact set[8] as in (9). In Section 4, we establish the convergence of $\widehat{\mathcal{A}}_n$ to $\mathcal{A}_+$ in probability under a certain metric with suitable assumptions. We develop our discussion under some higher level assumptions which should be justified for each example on hand. In the following, before going into the asymptotics of our estimator, we present semiparametric versions of two discrete games providing the inequality conditions of (2) which we are based on for our set estimation and inference.

3.1 Example1: Two Firms Entry-Exist Game

Suppose there are two potential entrants in a market whose profits depend on the existence of its rival. Let Π_j denote the profit of the firm $j = 1, 2$ as

$$\Pi_j(y_1, y_2|x) = \pi_j(Y_1 = y_1, Y_2 = y_2|X = (x_1', x_2')') + \varepsilon_j$$

where we let $\pi_1(1,0|x) = a + x_{1-c}'\theta_1 + h_1(x_{1c})$, $\pi_2(0,1) = a + x_{2-c}'\theta_2 + h_2(x_{2c})$, $\pi_1(1,1) = b + x_{1-c}'\theta_1 + h_1(x_{1c})$, $\pi_2(1,1) = b + x_{2-c}'\theta_2 + h_2(x_{2c})$, and $\pi_1(0,\cdot) = \pi_2(\cdot,0) = 0$ with $a > b$ since a monopoly profit tends to be higher than a duopoly profit. $\varepsilon = (\varepsilon_1, \varepsilon_2)$ are known to each firm but not to econometricians. Here the payoffs functions of players are given as partially linear forms. We let $X_{jc} \cap X_{j-c} = \varnothing$ and $X_{jc} \cup X_{j-c} = X_j$ for $j = 1, 2$ and let $h_1(0) = h_2(0)$ for the identification of parameters. Now assume that ε_j follows a normal distribution and $\varepsilon_1 \perp \varepsilon_2$. Then, the probability of being a monopolist will be $\Phi(\mu_{a_j} \equiv a + x_{j-c}'\theta_j + h_j(x_{jc})) = \Pr(\varepsilon_j > -\mu_{a_j})$ and that of being a duopolist will be $\Phi(\mu_{b_j} \equiv b + x_{j-c}'\theta_j + h_j(x_{jc})) = \Pr(\varepsilon_j > -\mu_{b_j})$. For this game, we note that multiple Nash equilibria exist depending on the realization of $(\varepsilon_1, \varepsilon_2)$ and that the necessary conditions of the Nash equilibria give us the following four inequality conditions comparing the true probabilities of events and the model probabilities as the form of (2):

$$\begin{aligned}
&P(0,0|\mu_{a_1}, \mu_{a_2}, \mu_{b_1}, \mu_{b_2}) = (1 - \Phi(\mu_{a_1}))(1 - \Phi(\mu_{a_2})) \geq P_0(0,0|x),\\
&P(0,1|\mu_{a_1}, \mu_{a_2}, \mu_{b_1}, \mu_{b_2}) = \Phi(\mu_{a_2})(1 - \Phi(\mu_{b_1})) \geq P_0(0,1|x),\\
&P(1,0|\mu_{a_1}, \mu_{a_2}, \mu_{b_1}, \mu_{b_2}) = \Phi(\mu_{a_1})(1 - \Phi(\mu_{b_2})) \geq P_0(1,0|x), \text{ and}\\
&P(1,1|\mu_{a_1}, \mu_{a_2}, \mu_{b_1}, \mu_{b_2}) = \Phi(\mu_{b_1})\Phi(\mu_{b_2}) \geq P_0(1,1|x).
\end{aligned}$$

This game model includes both finite dimensional parameters (a, b, and θ_j, $j = 1, 2$) and infinite dimensional parameters ($h_j(\cdot)$, $j = 1, 2$) of interest.

3.1.1 Construction of the Profiled Estimator

Now let $\theta = (a, b, \theta_1', \theta_2')'$ and $h = (h_1, h_2)$ and rewrite $P(Y_1, Y_2|X, \theta, h) = P(Y_1, Y_2|\mu_{a_1}, \mu_{a_2}, \mu_{b_1}, \mu_{b_2})$. Under the correct model specification with true parameters of (θ_0, h_0), we have $P(0,0|X, \theta_0, h_0) = P_0(0,0|X)$, $1 - P(0,0|X, \theta_0, h_0) - P(1,1|X, \theta_0, h_0) = P_0(0,1|X) + P_0(1,0|X)$, and $P(1,1|X, \theta_0, h_0) = P_0(1,1|X)$ regardless of the multiplicity of the Nash equilibria[9]. Using this fact and the method of sieve MLE[10] similarly with Kim (2006), we estimate h_0 as a profiled estimate of the form $\widehat{h}(\cdot, \theta)$ such that

$$\widehat{h}(\cdot, \theta) = \underset{h \in \mathcal{H}_n}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^{n} l(y_i, x_i; \theta, h)$$

[8] $Q_n(\theta, h)$ is continuous in θ as long as h is continuous in θ and $\widehat{c}_n(\cdot)$ is continuous in (θ, h), which is also continuous as long as $P(y|x, \theta, h)$ is continuous in (θ, h).

[9] Thus, we have redefined the outcome space in terms of the number of firms in the market.

[10] Alternatively, we can obtain a profiled estimator using the Sieve Minimum Distance estimator proposed by Ai and Chen (2003) noting that the model can be characterized in terms of the moment conditions: $0 = E\left[(1 - Y_1)(1 - Y_2) - P(0,0|X, \theta_0, h_0)|X\right]$ and $0 = E\left[Y_1 Y_2 - P(1,1|X, \theta_0, h_0)|X\right]$.

where $l(y_i, x_i; \theta, h) \equiv \mathbf{1}[y_{1i} + y_{2i} = 0] \log P(0,0|x_i,\theta,h) + \mathbf{1}[y_{1i} + y_{2i} = 1] \log(1 - P(0,0|x_i,\theta,h) - P(1,1|x_i, \theta, h)) + \mathbf{1}[y_{1i} + y_{2i} = 2] \log P(1,1|x_i,\theta,h)$. Under some regularity conditions similar with those in Kim (2006), we can show that $\sup_{\theta\in\Theta} \sup_{x_{jc}\in\mathbb{S}(X_{jc})} |\widehat{h}_j(\cdot,\theta) - h_{j0}(\cdot,\theta)| = o_p(1)$ for $j = 1, 2$. Interestingly, here we note that we may estimate the parameters simultaneously as $\left(\widehat{\theta}, \widehat{h}\right)$ such that

$$\left(\widehat{\theta}, \widehat{h}\right) = \underset{\theta\in\Theta, h\in\mathcal{H}_n}{\operatorname{argmax}} \ \tfrac{1}{n} \textstyle\sum_{i=1}^{n} l(y_i, x_i; \theta, h).$$

The consistency and the asymptotic normality of functional of parameters for such estimators can be also found in Kim (2006). However, we note that to achieve this consistent point estimation, we have to disregard the information about which firm enters the market since we redefine the outcome space in terms of how many firms in the market. Due to this omitted information, we may end up a point estimator whose variance is larger and thus we may have wider confidence interval for a parameter of interest than that of the set estimator using the full information in the model. It will be an interesting future research to compare these two alternatives: CI from the point estimation with the usage of less information vs. CI from the set estimation with the usage of the full information.

3.2 Example 2: Signaling Game with Two Discrete Types

The following example is a simple version of Kim (2006). Consider the *beer-quiche* game in Cho and Kreps (1987) where we have two players. Player 1 has either of two types $\{strong, weak\}$ with the probability of being the strong type equal to p_u,[11] and knows her type. After observing her type, Player 1 moves first sending one of two messages $\{Beer, Quiche\}$ to Player 2. Then, Player 2 chooses an action "*Fight*" or "*No Fight*" after observing the signal sent by Player 1. After the play, a payoff is realized depending on actions chosen by two players and Player 1's type. The structure and the payoffs of this game is given in Figure 1. Here φ_1 denotes the cost of mimicking the other type of Player 1 (cost of signalling *falsely*) and φ_2 measures Player 2's incentive to single out a particular type of Player 1. In the payoffs functions, Players can observe each other's ε but econometricians only know the distribution of ε such as normal distribution.

In this game we have four possible observable outcomes: Player 1 chooses beer but Player 2 decides to fight, Player 1 chooses beer and Player 2 decides not to fight, Player 1 chooses quiche and Player 2 decides to fight, or Player 1 chooses quiche but Player 2 decides not to fight. We let $Y_1 = 1$ if Player 1 chooses *Beer* and $Y_1 = 0$ otherwise and let $Y_2 = 1$ if Player 2 chooses *No Fight* and $Y_2 = 0$ otherwise. From the result of Kim (2006)[12], using the Perfect Bayesian Equilibrium (PBE), we can characterize the equilibrium of the game as summarized in Figure 2. As illustrated in the figure, depending on the realizations of $(\varepsilon_1, \varepsilon_2)$, we can have *Pooling equilibria*, *Separating equilibria*, or *Semi-separating equilibria*. Then, we obtain the model probabilities for each four possible outcome by integrating regions of ε for each particular observable outcome as below[13]. We let $\mu_1 \equiv X_1'\theta_1$, $\mu_2 \equiv X_2'\theta_2$, and $p \equiv p(Z)$ where $p(Z)$ is the conditional probability of being a strong type conditional on the public signal Z regarding Player 1's type, which are available to both Player 2 and econometricians.

[11] Note that p_u is the unconditional probability while $p(\cdot)$ denotes the conditional probability.

[12] When PBE is adopted as the equilibrium concept, Kim (2006) shows that this signaling game has multiple equilibria depending on the realizations of $(\varepsilon_1, \varepsilon_2)$. In the region $\mathcal{E}_1 \equiv \{(\varepsilon_1, \varepsilon_2) : \varepsilon_1 \geq \mu + X_1'\theta_1 + \varphi_1 \ \& \ X_2'\theta_2 + (2p-1)\varphi_2 \leq \varepsilon_2 \leq X_2'\theta_2 + \varphi_2\}$, we can have two equilibria: Pooling equilibrium with *Beer* & *Fight* or Pooling equilibrium with *Quiche* & *Fight* while in $\mathcal{E}_2 \equiv \{(\varepsilon_1, \varepsilon_2) : \varepsilon_1 \leq \mu + X_1'\theta_1 - \varphi_1 \ \& \ X_2'\theta_2 - \varphi_2 \leq \varepsilon_2 \leq X_2'\theta_2 + (2p-1)\varphi_2\}$, we can have two equilibria: Pooling with *Beer* & *No Fight* or Pooling with *Quiche* & *No Fight*. Kim (2006) also shows that we can still achieve the uniqueness of equilibrim by strenghtening the concept of equilibrim such as Cho and Kreps (1987)'s Intuitive Criterion. Only allowing equilibrium that survives this Intuitive Criterion, Kim (2006) shows that only (*Quiche*, *Fight*) survives in $\mathcal{E}_1$ and only (*Beer*, *No Fight*) survives in $\mathcal{E}_2$.

[13] For details how to derive the equilibria of the game and the resulting model probabilities, see Kim (2006).

$$
\begin{aligned}
&P\left(1,1|p,\mu_1,\mu_2,\varphi_1,\varphi_2\right)= \\
&\Phi\left(\mu_1-\varphi_1\right)\left(\Phi\left(\mu_2+(2p-1)\varphi_2\right)-\Phi\left(\mu_2-\varphi_2\right)\right)+p\Phi\left(\mu_2-\varphi_2\right) \\
&+p(\Phi\left(\mu_1+\varphi_1\right)-\Phi\left(\mu_1-\varphi_1\right))(\Phi\left(\mu_2+\varphi_2\right)-\Phi\left(\mu_2-\varphi_2\right)) \\
&+\int_0^1\phi\left(\mu_2+\left(\tfrac{p-\omega(1-p)}{p+\omega(1-p)}\right)\varphi_2\right)\tfrac{2p(1-p)\varphi_2}{p+(1-p)\omega}d\omega\int_0^1\phi\left(\mu_1-\tfrac{\varphi_1}{\omega}\right)\tfrac{\varphi_1}{\omega}d\omega \\
&+\int_0^1 p\omega\phi\left(\mu_2+\left(\tfrac{(1-\omega)p-(1-p)}{(1-\omega)p+(1-p)}\right)\varphi_2\right)\tfrac{2p(1-p)\varphi_2}{(1-\omega p)^2}d\omega\int_0^1\omega\phi\left(\mu_1+\tfrac{\varphi_1}{1-\omega}\right)\tfrac{\varphi_1}{(1-\omega)^2}d\omega
\end{aligned}
$$

$$
\begin{aligned}
&P\left(1,0|p,\mu_1,\mu_2,\varphi_1,\varphi_2\right)= \\
&\left(1-\Phi\left(\mu_1+\varphi_1\right)\right)\left(\Phi\left(\mu_2+\varphi_2\right)-\Phi\left(\mu_2+(2p-1)\varphi_2\right)\right)+p\left(1-\Phi\left(\mu_2+\varphi_2\right)\right) \\
&+\int_0^1\phi\left(\mu_2+\left(\tfrac{p-\omega(1-p)}{p+\omega(1-p)}\right)\varphi_2\right)\tfrac{2p(1-p)\varphi_2}{p+(1-p)\omega}d\omega\int_0^1\left(1-\omega\right)\phi\left(\mu_1-\tfrac{\varphi_1}{\omega}\right)\tfrac{\varphi_1}{\omega^2}d\omega \\
&+\int_0^1 p\omega\phi\left(\mu_2+\left(\tfrac{(1-\omega)p-(1-p)}{(1-\omega)p+(1-p)}\right)\varphi_2\right)\tfrac{2p(1-p)\varphi_2}{(1-\omega p)^2}d\omega\int_0^1\phi\left(\mu_1+\tfrac{\varphi_1}{1-\omega}\right)\tfrac{\varphi_1}{1-\omega}d\omega
\end{aligned}
$$

$$
\begin{aligned}
&P\left(0,1|p,\mu_1,\mu_2,\varphi_1,\varphi_2\right)= \\
&\Phi\left(\mu_1-\varphi_1\right)\left(\Phi\left(\mu_2+(2p-1)\varphi_2\right)-\Phi\left(\mu_2-\varphi_2\right)\right)+\left(1-p\right)\Phi\left(\mu_2-\varphi_2\right) \\
&+\int_0^1\left(1-p\right)\left(1-\omega\right)\phi\left(\mu_2+\left(\tfrac{p-\omega(1-p)}{p+\omega(1-p)}\right)\varphi_2\right)\tfrac{2p(1-p)\varphi_2}{(p+\omega(1-p))^2}d\omega\int_0^1\phi\left(\mu_1-\tfrac{\varphi_1}{\omega}\right)\tfrac{\varphi_1}{\omega}d\omega \\
&+\int_0^1\phi\left(\mu_2+\left(\tfrac{(1-\omega)p-(1-p)}{(1-\omega)p+(1-p)}\right)\varphi_2\right)\tfrac{2p(1-p)\varphi_2}{1-\omega p}d\omega\int_0^1\omega\phi\left(\mu_1+\tfrac{\varphi_1}{1-\omega}\right)\tfrac{\varphi_1}{(1-\omega)^2}d\omega
\end{aligned}
$$

$$
\begin{aligned}
&P\left(0,0|p,\mu_1,\mu_2,\varphi_1,\varphi_2\right)= \\
&(1-\Phi\left(\mu_1+\varphi_1\right))(\Phi(\mu_2+\varphi_2)-\Phi(\mu_2+(2p-1)\varphi_2))+\left(1-p\right)\left(1-\Phi\left(\mu_2+\varphi_2\right)\right) \\
&+\left(1-p\right)\left(\left(\Phi\left(\mu_1+\varphi_1\right)-\Phi\left(\mu_1-\varphi_1\right)\right)\left(\Phi\left(\mu_2+\varphi_2\right)-\Phi\left(\mu_2-\varphi_2\right)\right)\right) \\
&+\int_0^1\left(1-p\right)\left(1-\omega\right)\phi\left(\mu_2+\left(\tfrac{p-\omega(1-p)}{p+\omega(1-p)}\right)\varphi_2\right)\tfrac{2p((1-p)\varphi_2}{(p+\omega(1-p))^2}d\omega\int_0^1\left(1-\omega\right)\phi\left(\mu_1-\tfrac{\varphi_1}{\omega}\right)\tfrac{\varphi_1}{\omega^2}d\omega \\
&+\int_0^1\phi\left(\mu_2+\left(\tfrac{(1-\omega)p-(1-p)}{(1-\omega)p+(1-p)}\right)\varphi_2\right)\tfrac{2p(1-p)\varphi_2}{1-\omega p}d\omega\int_0^1\phi\left(\mu_1+\tfrac{\varphi_1}{1-\omega}\right)\tfrac{\varphi_1}{1-\omega}d\omega.
\end{aligned}
$$

This game model also includes both finite dimensional parameters (φ_j, p_u, and β_j, $j=1,2$) and infinite dimensional parameters ($p(\cdot)$) of interest. The estimation and inference of this signaling game model will be based on the four inequality conditions as the form of (2): $P\left(y_1,y_2|p,\mu_1,\mu_2,\varphi_1,\varphi_2\right)\geq P_0\left(y_1,y_2|X=x\right)$ where $X=X_1\cup X_2\cup Z$ and y_1, $y_2=\{0,1\}$.

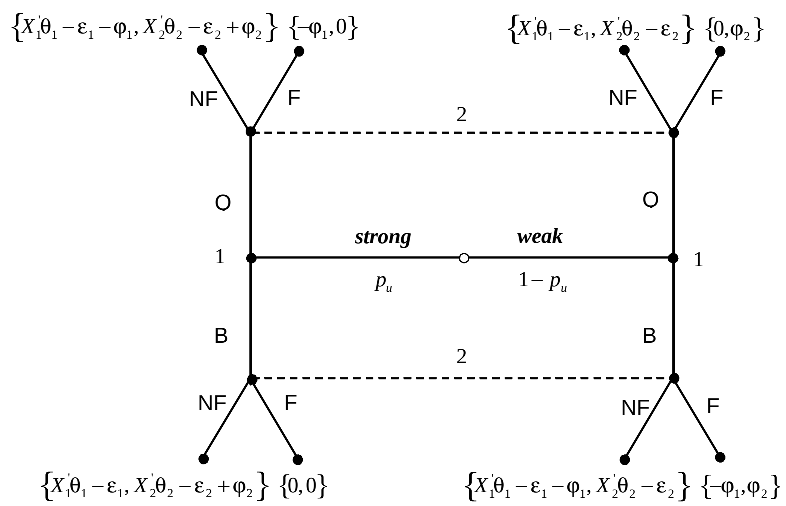

Figure 1: Signaling Game with Two Discrete Types

3.2.1 Construction of the Profiled Estimator

Now let[14] $\theta = (\varphi_1, \varphi_2, \theta_1', \theta_2')'$ and $h = L^{-1}(p(Z))$ with $L(\cdot) = \exp(\cdot)/(1+\exp(\cdot))$ under the logit specification of $p(\cdot)$. Rewrite $P(Y_1, Y_2|X, \theta, h) = P(Y_1, Y_2|p(\cdot), \mu_1, \mu_2, \varphi_1, \varphi_2)$. To obtain a consistent estimate of $h_0(\cdot, \theta)$, we again consider a redefinition of outcome space. From Figure 2, we note that regardless of the multiple equilibria in the regions of $\mathcal{E}_1 \equiv \{(\varepsilon_1, \varepsilon_2) : \varepsilon_1 \geq \mu + X_1'\theta_1 + \varphi_1 \;\&\; X_2'\theta_2 + (2p-1)\varphi_2 \leq \varepsilon_2 \leq X_2'\theta_2 + \varphi_2\}$ and $\mathcal{E}_2 \equiv \{(\varepsilon_1, \varepsilon_2) : \varepsilon_1 \leq \mu + X_1'\theta_1 - \varphi_1 \;\&\; X_2'\theta_2 - \varphi_2 \leq \varepsilon_2 \leq X_2'\theta_2 + (2p-1)\varphi_2\}$, we have a well-defined likelihood function when we redefine the outcomes of the game in terms of whether a fight is raised or not. Figure 3[15] shows the resulting redefinition of outcome space. From this observation, we have

$$
\begin{aligned}
P(Y_2 &= 0|X, \theta, h) = P(0,0|X, \theta, h) + P(1,0|X, \theta, h) - \Pr(\varepsilon \in \mathcal{E}_1) \\
P(Y_2 &= 1|X, \theta, h) = P(0,1|X, \theta, h) + P(1,1|X, \theta, h) - \Pr(\varepsilon \in \mathcal{E}_2)
\end{aligned}
$$

where $\Pr(\varepsilon \in \mathcal{E}_1) = (1 - \Phi(\mu_1 + \varphi_1))(\Phi(\mu_2 + \varphi_2) - \Phi(\mu_2 + (2p-1)\varphi_2))$ and $\Pr(\varepsilon \in \mathcal{E}_2) = \Phi(\mu_1 - \varphi_1)(\Phi(\mu_2 + (2p-1)\varphi_2) - \Phi(\mu_2 - \varphi_2))$. Then, we obtain a consistent profiled estimator, $\widehat{h}(\cdot, \theta)$, using the sieve MLE similarly with Example 1 such that

$$
\widehat{h}(\cdot, \theta) = \underset{h \in \mathcal{H}_n}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^n l(y_i, x_i; \theta, h) \equiv \{\mathbf{1}[y_{2i} = 1] \log P(0|x_i, \theta, h) + \mathbf{1}[y_{2i} = 0] \log P(1|x_i, \theta, h)\}
$$

Under some regularity conditions similar with those in Kim (2006), we can also show that $\sup_{\theta \in \Theta} \sup_{z \in \mathbb{S}(z)} \left|\widehat{h} - h_0\right| = o_p(1)$. As in Example 1, here we can also estimate the parameters simultaneously as $\left(\widehat{\theta}, \widehat{h}\right)$ such that $\left(\widehat{\theta}, \widehat{h}\right) = \underset{\theta \in \Theta, h \in \mathcal{H}_n}{\operatorname{argmax}} \frac{1}{n} \sum_{i=1}^n l(y_i, x_i; \theta, h)$. However, this consistent point estimation again requires the information reduction. We do not use the information regarding Player 1's actions. Because of this, we may obtain a point estimator with larger variance and thus we may have wider confidence interval than that of the set estimator considered in this paper. This concern justifies the use of the set inference even though we can achieve a consistent point estimation.

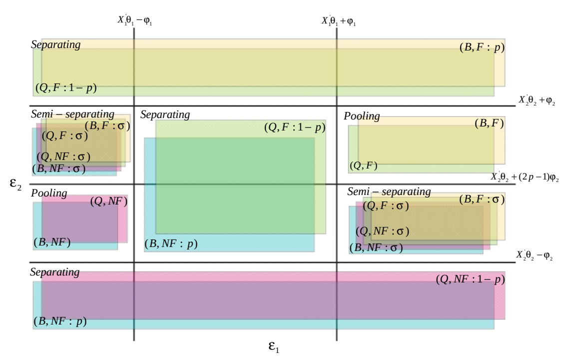

Figure 2: Equilibrium of the Game

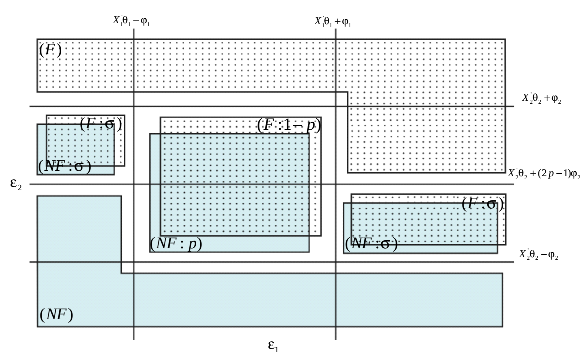

Figure 3: Redefinition of Outcome Space

[14] We note that in the conditional probabilities of observed outcomes, implied by the model, p_u does not appear separately from $p(Z)$. Nontheless, we can still estimate p_u as a functional of $p(Z)$. We will discuss this in Section 4.1.

[15] The dotted area is for "*Fight*" and the solid area is for "*No Fight*". "σ" denotes the Semi-Separating Equilibria

4 Probability Limit of the Set Estimator

We use a version of the Hausdorff metric measuring the distance between two sets whose elements are (e_1, e_2)'s such that $e_1 \in \Theta$ and $e_2 = (e_{21}, \ldots, e_{2\boldsymbol{v}})$ belongs to a class of vector of continuous functions defined on $\mathbb{S}(Z_1) \times \ldots \times \mathbb{S}(Z_{\boldsymbol{v}})$ where Z_v's are subsets of X. For two such sets A and B whose elements are $a = (a_1, a_2)$ and $b = (b_1, b_2)$ respectively, let the maximum distance between any points in A and B be given by $\rho(A|B) = \sup_{a \in A} \rho(a|B)$, where $\rho(a|B) = \inf \{\|a - b\|_s = \|a_1 - b_1\|_E + \|a_2 - b_2\|_{\mathcal{H}} : b \in B\}$ for $a \in A$ and a pseudo metric $\|\cdot\|_{\mathcal{H}}$ is defined by $\|\cdot\|_{\mathcal{H}} \equiv \sup_{\|a_1' - b_1'\|_{\acute{E}} \leq \|a_1 - b_1\|_E, a_1', b_1' \in \Theta} \left\{ \sum_{v=1}^{\boldsymbol{v}} \sup_{z_v \in \mathbb{S}(Z_v)} |a_{2v}(z_v) - b_{2v}(z_v)| \right\}$. Notice that $\rho(a|B)$ is the distance from an element a to the set B. By definition, if $A = \varnothing$ and $B \neq \varnothing$, then $\rho(A|B) = 0$ while $\rho(A|B) = \infty$ if $B = \varnothing$. The Hausdorff metric distance between A and B is given by $d(A, B) = \max\{\rho(A|B), \rho(B|A)\}$. For completeness, we also let $\|a - b\|_s = \|a_1 - b_1\|_E$ and define the Hausdorff metric accordingly when we have only the finite dimensional parameters. Also note that $\|\cdot\|_{\mathcal{H}} = \sup_{\theta \in \Theta} \left\{ \sum_{v=1}^{\boldsymbol{v}} \sup_{z_v \in \mathbb{S}(Z_v)} |a_{2v}(z_v) - b_{2v}(z_v)| \right\}$ when $a_1 = b_1$. Here we provide some conditions under which

$$\text{(i) } \rho(\widehat{\mathcal{A}}_n | \mathcal{A}_+) \underset{p}{\rightarrow} 0 \text{ and (ii) } \rho(\mathcal{A}_+ | \widehat{\mathcal{A}}_n) \underset{p}{\rightarrow} 0 \tag{11}$$

and thus we establish the conditions for $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \underset{p}{\rightarrow} 0$. Note that (11)(i) ensures that $\widehat{\mathcal{A}}_n$ is not larger than $\mathcal{A}_+$ asymptotically while (11)(ii) ensures that $\widehat{\mathcal{A}}_n$ is not smaller than $\mathcal{A}_+$ asymptotically. Also note that (11)(ii) alone ensures that the distance of the true value α_0 $(\in \mathcal{A}_0 \subset \mathcal{A}_+)$ from the set estimator $\widehat{\mathcal{A}}_n$ satisfies $\rho(\alpha_0 | \widehat{\mathcal{A}}_n) \underset{p}{\rightarrow} 0$.

Let Γ_{all} be a class of subsets of $\mathbb{S}(X)$ that includes all possible realizations of $\widehat{\gamma}_{n,j,m}$ for all $(j, m) \in \mathcal{I}_{J,\mathcal{M}}$ and $n \geq 1$. Also let $\mathcal{I}_J = \{1, \ldots, J\}$. We use "wpa1" to denote "with probability that approaches to one". The following assumptions are sufficient to establish the result (11)(i).

Assumption 4.1 *$\{(Y_i, X_i)\}_{i=1}^n$ are iid.*

Assumption 4.2 *The true parameter α_0 satisfies $P(y|x, \alpha_0) - P_0(y|x) \geq 0$, $\forall (y, x) \in \mathbb{S}(Y) \times \mathbb{S}(X)$.*

Assumption 4.3 *(i) $\Theta \times \mathcal{H}$ is compact under the metric $\|\cdot\|_s$; (ii) $\mathcal{H}_n \subseteq \mathcal{H}_{n+1} \subseteq \mathcal{H}$ for all $n \geq 1$ and for any $h \in \mathcal{H}$, there exists $\Pi_n h \in \mathcal{H}_n$ such that $\|\Pi_n h - h\|_{\mathcal{H}} = o(1)$; (iii) $\left\|\widehat{h} - h_0\right\|_{\mathcal{H}} = o_p(1)$; (iv) $h(\cdot, \theta)$ is continuous in θ for all h s.t. $\|h - h_0\|_{\mathcal{H}} = o(1)$.*

Assumption 4.4 *(i) $P(y|x, \alpha)$ is Hölder continuous in α on $\mathcal{A}$; (ii) $|q_\gamma(x)| < \infty$ for all $\gamma \in \Gamma_{all}$ uniformly over $x \in \mathbb{S}(X)$.*

To establish the second result (11) (ii), we need additional assumptions. The following conditions are in the line with ABJ. Let int(A) and cl(A) denote the interior and closure of a set A, respectively.

Assumption 4.5 *Either (i) $\Theta_+ = \{\theta_0\}$ or*
(ii) (a) $\Theta_+ =$cl(int(Θ_+)) and (b) $\forall \theta \in$int(Θ_+), $\inf_{(j,m) \epsilon \mathcal{I}_{J,\mathcal{M}}} c_0\left(j, \gamma_{0,j,m}, \theta, h_0(\cdot, \theta)\right) > 0$.

Assumption 4.2 states that the model (conditional probabilities implied by the game) is correctly specified, which ensures that $\mathcal{A}_0$ and $\mathcal{A}_+$ are not empty. Note that Assumptions 4.3 and 4.4 are standard assumptions in the semi-nonparametric literature. Note that Assumption 4.5 (i) holds when the necessary conditions (5) are strong enough that $\mathcal{A}_+$ only contains the true parameter $\alpha_0 = (\theta_0, h_0)$. Assumption 4.5 (ii) (a) implies that Θ_+ has a non-empty interior and does not contain isolated points. Assumption 4.5 (ii) (b) requires that for any $\theta \in$int(Θ_+), the necessary conditions (5) hold with a strict inequality. We also need some consistency results for $\widehat{c}_n(\cdot)$ and $\widehat{\Gamma}_n$ under certain metrics. First, define $\mathcal{H}_{\delta_n} \equiv \{h \in \mathcal{H} : \|h(\cdot,\theta) - h_0(\cdot,\theta)\|_{\mathcal{H}} \leq \delta_n,\ \theta \in \Theta\}$, $\mathcal{H}_{n,\delta_n} \equiv \{h \in \mathcal{H}_n : \|h(\cdot,\theta) - h_0(\cdot,\theta)\|_{\mathcal{H}} \leq \delta_n,\ \theta \in \Theta\}$ with $\delta_n = o(1)$, and for any two real functions c_1 and c_2 on $\mathcal{I}_J \times \Gamma_{all} \times \Theta \times \mathcal{H}_{\delta_n}$, define $\|c_1 - c_2\|_{U_n} \equiv \sup_{(j,\gamma,\theta,h)\in\mathcal{I}_J\times\Gamma_{all}\times\Theta\times\mathcal{H}_{\delta_n}} |c_1(j,\gamma,\theta,h) - c_2(j,\gamma,\theta,h)|$. The following is the semiparametric version of Assumption 5 in ABJ.

Assumption 4.6 $\|\widehat{c}_n(\cdot) - c_0(\cdot)\|_{U_n} \underset{p}{\rightarrow} 0.$

Let $\mathcal{F}_\xi = \left\{\xi\left(y,x,j,\gamma,\theta,h\right) = \left(P(y^{(j)}|x,\theta,h) - \mathbf{1}[y = y^{(j)}]\right) q_\gamma(x) : (j,\gamma,\alpha) \in \mathcal{I}_J \times \Gamma_{all} \times \Theta \times \mathcal{H}\right\}$ denote the class of measurable functions indexed by (j,γ,θ,h). Assumption 4.6 will hold when $\mathcal{F}_\xi$ is a P-Glivenko-Cantelli class as presented in van der Vaart and Wellner (1996). Now define a semi-norm $\|\cdot\|$ as follows. For γ_1 and $\gamma_2 \in \Gamma_{all}$, we let

$$\|\gamma_1 - \gamma_2\| \equiv \left(\int \left|q_{\gamma_1}(x) - q_{\gamma_2}(x)\right|^2 dF_X(x)\right)^{1/2} \text{ and } \|\Gamma_1 - \Gamma_2\| \equiv \max_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} \left\|\gamma_{1,j,m} - \gamma_{2,j,m}\right\| \tag{12}$$

where $\Gamma_1 = \left\{\gamma_{1,j,m} \in \Gamma_{all} : (j,m) \in \mathcal{I}_{J,\mathcal{M}}\right\}$ and $\Gamma_2 = \left\{\gamma_{2,j,m} \in \Gamma_{all} : (j,m) \in \mathcal{I}_{J,\mathcal{M}}\right\}$. We assume

Assumption 4.7 $\left\|\widehat{\Gamma}_n - \Gamma_0\right\| \underset{p}{\rightarrow} 0$ *(Assumption 6 of ABJ).*

Now we are ready to present the consistency result of the set estimator $\widehat{\mathcal{A}}_n$.

Theorem 4.1 *(i) Suppose Assumptions 4.1, 4.2, 4.3, 4.4, 4.6, and 4.7 hold. Then,* $\rho(\widehat{\Theta}_n|\Theta_+) \underset{p}{\rightarrow} 0$.
(ii) Under Assumptions 4.3, 4.4, 4.5, 4.6,and 4.7, $\rho(\Theta_+|\widehat{\Theta}_n) \underset{p}{\rightarrow} 0$. *Thus, we have* $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \underset{p}{\rightarrow} 0$.

In Section 5, we discuss how to construct the confidence interval of a real functional $\beta_n(\alpha)$ of α. In particular, we may restrict our interests to real functions such that $\beta_n(\alpha) = r_n(\theta)$ where $r_n(\theta)$ is a real function of θ. For example, we can have $\beta_n(\alpha) = \theta_{(k)}$ where $\theta_{(k)}$ is the k-th element of θ. When constructing the confidence interval of the real functional $\beta_n(\alpha)$, its largest and smallest values are of interest. The largest and smallest values of $\beta_n(\alpha)$ across all $\alpha \in \mathcal{A}_+$ defined by

$$\beta_{n,U} = \sup\left\{\beta_n(\alpha) : \alpha \in \mathcal{A}_+\right\} \text{ and } \beta_{n,L} = \inf\left\{\beta_n(\alpha) : \alpha \in \mathcal{A}_+\right\}, \tag{13}$$

respectively. We estimate these values, respectively, by

$$\widehat{\beta}_{n,U} = \sup\left\{\beta_n(\alpha) : \alpha \in \widehat{\mathcal{A}}_n\right\} \text{ and } \widehat{\beta}_{n,L} = \inf\left\{\beta_n(\alpha) : \alpha \in \widehat{\mathcal{A}}_n\right\}. \tag{14}$$

The consistency of $\widehat{\beta}_{n,U}$ for $\beta_{n,U}$ and $\widehat{\beta}_{n,L}$ for $\beta_{n,L}$ is obtained from Corollary 4.1 as long as $\beta_n(\cdot)$ has some continuity property (with respect to the metric $\|\cdot\|_s$). We note that if $\{\beta_n(\cdot) : n \geq 1\}$ is stochastically equicontinuous on $\mathcal{A}$, the consistency results hold.

Assumption 4.8 $\{\beta_n(\cdot) : n \geq 1\}$ *is stochastically equicontinuous on* $\mathcal{A}$*, i.e. for any given* $\varepsilon > 0$ *and any* $\alpha \in \mathcal{A}$*, there exists* $\delta > 0$ *such that* $\limsup_{n\rightarrow\infty} P\left(\sup_{\|\widetilde{\alpha}-\alpha\|_s\leq\delta} |\beta_n(\widetilde{\alpha}) - \beta_n(\alpha)| > \varepsilon\right) < \varepsilon$.

In particular, when $\beta_n(\cdot)$ is (pointwise) Lipschitz continuous with respect to α, primitive sufficient conditions for stochastic equicontinuity can be found in Andrews (1994) or Newey and McFadden (1994). Even when $\beta_n(\cdot)$ is (pointwise) Lipschitz continuous with respect to h but not in θ, we can still apply the results in Andrews (1994) for certain cases. Chen, Linton, and van Keilegom (2003) also provide some stochastic equicontinuity results even when $\beta_n(\cdot)$ is not (pointwise) continuous with respect to h and θ. From the result of Theorem 4.1, $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \underset{p}{\rightarrow} 0$, we obtain

Corollary 4.1 *Under Assumptions 4.1-4.8,* $\widehat{\beta}_{n,U} - \beta_{n,U} \underset{p}{\rightarrow} 0$ *and* $\widehat{\beta}_{n,L} - \beta_{n,L} \underset{p}{\rightarrow} 0$.

4.1 Example: Set Estimation of the Type Distribution

We note that in the conditional probabilities of observed outcomes, implied by the model, p_u does not appear separately from $p(Z)$ in Section 3.2 as originally noted in Kim (2006). However, using the law of iterated expectation $p_u = E[p(Z)]$, we can still identify the type distribution parameter p_u. Recalling $p(Z) = L(h(Z)) \equiv \exp(h(Z))/(1+\exp(h(Z)))$, we obtain a set estimator of p_u such that

$$\widehat{\mathcal{P}}_n = \left\{p_u : p_u = \tfrac{1}{n}\textstyle\sum_{i=1}^n L(h(z_i)) = \tfrac{1}{n}\sum_{i=1}^n \tfrac{\exp(h(z_i))}{1+\exp(h(z_i))} \text{ for each } h \in \widehat{\mathcal{H}}_n\right\}.$$

We note that as long as $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) = o_p(1)$, $\widehat{\mathcal{P}}_n$ converges to its population counterpart $\mathcal{P}_+$ defined by $\mathcal{P}_+ = \{p_u : p_u = E[L(h)] = E\left[\frac{\exp(h)}{1+\exp(h)}\right] \text{ for each } h \in \mathcal{H}_+\}$.

Proposition 4.1 *Suppose $\{Z_i\}_{i=1}^n$ are iid and $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) = o_p(1)$. Then, $d(\widehat{\mathcal{P}}_n, \mathcal{P}_+) = o_p(1)$.*

5 Confidence Intervals

In this section, we construct a CI[16] for the true value $\beta_0 = \beta_n(\alpha_0)$, where $\beta_n(\cdot)$ is a known real functional of α_0. Note that we suppress the potential dependence of β_0 on n for notational simplicity. In the spirit of Imbens and Manski (2003), the CI we consider is for the true value β_0, not for the set values $\beta_n(\alpha)$ for $\alpha \in \mathcal{A}_+$. The CI provided here is an extension of ABJ to the semiparametric case. Now let

$$\widehat{\mathcal{A}}_{n,U} = \{\alpha \in \widehat{\mathcal{A}}_n : \beta_n(\alpha) = \widehat{\beta}_{n,U}\}. \tag{15}$$

Note that $\widehat{\mathcal{A}}_{n,U}$ is not empty since $\widehat{\mathcal{A}}_n$ is compact under a metric $\|\cdot\|_s$. The compactness of $\widehat{\mathcal{A}}_n$ under $\|\cdot\|_s$ comes from (i) $\mathcal{A}_n$ is compact under $\|\cdot\|_s$ and (ii) $\widehat{\mathcal{A}}_n$ is defined using the non-strict inequality. We choose a unique value $\widehat{\alpha}_{n,U}$ such that

$$\widehat{\alpha}_{n,U} = \operatorname{argmin}\left\{\|\alpha\|_s : \alpha \in \widehat{\mathcal{A}}_{n,U}\right\}. \tag{16}$$

Again note that the existence of $\widehat{\alpha}_{n,U}$ is guaranteed since $\widehat{\mathcal{A}}_{n,U}$ is compact under $\|\cdot\|_s$. The solution of (16) may not be unique. In that case, a researcher can choose a particular value of $\widehat{\alpha}_{n,U}$ according to certain criterion. We define $\widehat{\mathcal{A}}_{n,L}$ and $\widehat{\alpha}_{n,L}$ analogously replacing $\widehat{\beta}_{n,U}$ with $\widehat{\beta}_{n,L}$. Note that by construction, we have $\beta_n(\widehat{\alpha}_{n,U}) = \widehat{\beta}_{n,U}$ and $\beta_n(\widehat{\alpha}_{n,L}) = \widehat{\beta}_{n,L}$. Now let $\widehat{\mathcal{B}}_{n,U}$ and $\widehat{\mathcal{B}}_{n,L}$ be collections of $(j,m) \in \mathcal{I}_{J,\mathcal{M}}$ such that the corresponding constraints bind at $\widehat{\alpha}_{n,U}$ and $\widehat{\alpha}_{n,L}$, respectively, which are the boundary points of $\widehat{\mathcal{A}}_n$. Thus, we have $\widehat{\mathcal{B}}_{n,U} = \left\{(j,m) \in \mathcal{I}_{J,\mathcal{M}} : \widehat{c}_n(j, \widehat{\gamma}_{n,j,m}, \widehat{\alpha}_{n,U}) = 0\right\}$ and $\widehat{\mathcal{B}}_{n,L} = \left\{(j,m) \in \mathcal{I}_{J,\mathcal{M}} : \widehat{c}_n(j, \widehat{\gamma}_{n,j,m}, \widehat{\alpha}_{n,L}) = 0\right\}$.

Now we are ready to present the $(1-a)$-CI of the true value β_0. We consider several versions of CI's depending on the choice of the sieve spaces that are used for constructing upper and lower bounds of CI's. We have not considered a specific function space and a sieve space for the estimation stage but in the inference stage, we need to do so since the construction of CI's critically depend on the choice of the function space for h. A Hölder space, denoted by $\Lambda^\nu(\mathbb{S}(Z))$, is a space of functions $g : \mathbb{S}(Z) \to \mathbb{R}$ such that the first $\underline{\nu}$ derivatives are bounded, and the $\underline{\nu}$-th derivatives are Hölder continuous with the exponent $\nu - \underline{\nu} \in (0,1]$, where $\underline{\nu}$ is the largest integer smaller than ν. The Hölder space becomes a Banach space when endowed with the Hölder norm: $||g||_{\Lambda^\nu} = \sup_z |g(z)| + \max_{a_1+a_2+\ldots a_{d_z} = \underline{\nu}} \sup_{z \neq z'} \frac{|\nabla^a g(z) - \nabla^a g(z')|}{(||z-z'||_E)^{\nu-\underline{\nu}}} < \infty$, where $\nabla^a g(z) \equiv \frac{\partial^{a_1+a_2+\ldots a_{d_z}}}{\partial z_1^{a_1} \ldots \partial z_{d_z}^{a_{d_z}}} g(z)$. The Hölder ball (with radius c) $\Lambda_c^\nu(\mathbb{S}(Z))$ is defined accordingly as $\Lambda_c^\nu(\mathbb{S}(Z)) \equiv \{g \in \Lambda^\nu(\mathbb{S}(Z)) : ||g||_{\Lambda^\nu} \le c < \infty\}$.

[16] We may consider the CI for identified set of $\beta_n(\cdot)$ alternatively. Such CI can be found in ABJ. The asymptotic justification of the semiparametric version of this can be given similary with Theorem 5.1. In this paper, we provide the CI for the true value only since our focus here is in extending ABJ to the semiparametric models.

Now let $\mathcal{H} \equiv \mathcal{H}^1 \times \ldots \times \mathcal{H}^{\boldsymbol{v}} = \Lambda_{c_1}^{\nu_1}(\mathbb{S}(Z_1)) \times \ldots \times \Lambda_{c_{\boldsymbol{v}}}^{\nu_{\boldsymbol{v}}}(\mathbb{S}(Z_{\boldsymbol{v}}))$. Then, it is well known that functions in $\mathcal{H}$ can be approximated well by power series, Fourier series, splines, and wavelets[17]. For example, we may let $\mathcal{H}^1 = \{h_1 : h_1(z_1) = \sum_{k=1}^{\infty} [a_k \cos(kz_1) + b_k \sin(kz_1)], ||h_1||_{\Lambda^{\nu_1}} \leq c_1\}$ where $h_1(z_1)$ is given as an infinite Fourier series and its derivative with a fractional power is also defined in terms of Fourier series.

The $(1-a)$-CI of the true value β_0 is given by

$$CI_n(1-a) = [\widetilde{\beta}_{n,L}, \widetilde{\beta}_{n,U}] \tag{17}$$

for some upper and lower bounds, $\widetilde{\beta}_{n,U}$ and $\widetilde{\beta}_{n,L}$ such that

$$\liminf_{n\to\infty} P(\beta_0 \subset CI_n(1-a)) = \liminf_{n\to\infty} P(\widetilde{\beta}_{n,L} \leq \beta_0 \leq \widetilde{\beta}_{n,U}) \geq 1-a. \tag{18}$$

We will consider three alternatives. First, define $\widehat{\mathcal{H}}_{\delta_n} \equiv \{h \in \mathcal{H} : \left\| h - \widehat{h}(\cdot,\theta) \right\|_{\mathcal{H}} \leq \delta_n,\ \theta \in \Theta\}$ and $\widehat{\mathcal{H}}_{l,\delta_n} \equiv \{h \in \mathcal{H}_l : \left\| h - \widehat{h}(\cdot,\theta) \right\|_{\mathcal{H}} \leq \delta_n,\ \theta \in \Theta\}$[18] with $\delta_n = o(1)$ where $\mathcal{H}_l$ is a finite dimensional sieve space such that $\mathcal{H}_l \subseteq \mathcal{H}_{l+1} \subseteq \mathcal{H}$ for all $l \geq 1$. Then, the upper and lower bounds, $\widetilde{\beta}_{n,U}$ and $\widetilde{\beta}_{n,L}$ for three alternative CI's are given as the following form:

- Alternative CI1: CI over the whole infinite dimensional space $(\Theta \times \mathcal{H})$: $\widetilde{\beta}_{n,U} \equiv \widetilde{\beta}_{n,U}^{(1)}$ and $\widetilde{\beta}_{n,L} \equiv \widetilde{\beta}_{n,L}^{(1)}$ such that

$$\begin{aligned} \widetilde{\beta}_{n,U}^{(1)} &= \sup\left\{\beta_n(\alpha) : \alpha \in \Theta \times \mathcal{H}, \widetilde{c}_{n,U}(j, \widehat{\gamma}_{n,j,m}, \alpha) \geq 0,\ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}\right\} \\ \widetilde{\beta}_{n,L}^{(1)} &= \inf\left\{\beta_n(\alpha) : \alpha \in \Theta \times \mathcal{H}, \widetilde{c}_{n,L}(j, \widehat{\gamma}_{n,j,m}, \alpha) \geq 0,\ \forall (j,m) \in \widehat{\mathcal{B}}_{n,L}\right\}, \end{aligned} \tag{19}$$

- Alternative CI2: CI over the infinite dimensional space around the true value h_0 $(\Theta \times \widehat{\mathcal{H}}_{\delta_n})$: $\widetilde{\beta}_{n,U} \equiv \widetilde{\beta}_{n,U}^{(2)}$ and $\widetilde{\beta}_{n,L} \equiv \widetilde{\beta}_{n,L}^{(2)}$ such that

$$\begin{aligned} \widetilde{\beta}_{n,U}^{(2)} &= \sup\left\{\beta_n(\alpha) : \alpha \in \Theta \times \widehat{\mathcal{H}}_{\delta_n}, \widetilde{c}_{n,U}(j, \widehat{\gamma}_{n,j,m}, \alpha) \geq 0,\ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}\right\} \\ \widetilde{\beta}_{n,L}^{(2)} &= \inf\left\{\beta_n(\alpha) : \alpha \in \Theta \times \widehat{\mathcal{H}}_{\delta_n}, \widetilde{c}_{n,L}(j, \widehat{\gamma}_{n,j,m}, \alpha) \geq 0,\ \forall (j,m) \in \widehat{\mathcal{B}}_{n,L}\right\}, \end{aligned} \tag{20}$$

- Alternative CI3: CI over the finite dimensional sieve space around the true value h_0 $(\Theta \times \widehat{\mathcal{H}}_{l,\delta_n})$: $\widetilde{\beta}_{n,U} \equiv \widetilde{\beta}_{l,n,U}^{(3)}$ and $\widetilde{\beta}_{n,L} \equiv \widetilde{\beta}_{l,n,L}^{(3)}$ such that

$$\begin{aligned} \widetilde{\beta}_{l,n,U}^{(3)} &= \sup\left\{\beta_n(\alpha) : \alpha \in \Theta \times \widehat{\mathcal{H}}_{l,\delta_n}, \widetilde{c}_{n,U}(j, \widehat{\gamma}_{n,j,m}, \alpha) \geq 0,\ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}\right\} \\ \widetilde{\beta}_{l,n,L}^{(3)} &= \inf\left\{\beta_n(\alpha) : \alpha \in \Theta \times \widehat{\mathcal{H}}_{l,\delta_n}, \widetilde{c}_{n,L}(j, \widehat{\gamma}_{n,j,m}, \alpha) \geq 0,\ \forall (j,m) \in \widehat{\mathcal{B}}_{n,L}\right\}, \end{aligned} \tag{21}$$

For Alternative CI3, we require $l \geq n$ so that $\widehat{\mathcal{H}}_{l,\delta_n} \supseteq \widehat{\mathcal{H}}$. This guarantees that $\widetilde{\beta}_{l,n,U}^{(3)} \geq \widehat{\beta}_{n,U}$ and $\widetilde{\beta}_{l,n,L}^{(3)} \leq \widehat{\beta}_{n,L}$, which are necessary to justify the proposed CI asymptotically. Here $\widetilde{c}_{n,U}$ is an upper bound on c_0 for those $\left(y^{(j)}, \widehat{\gamma}_{n,j,m}\right)$ sets for which (j,m) belongs to $\widehat{\mathcal{B}}_{n,U}$ at a particular value $\alpha = \widehat{\alpha}_{n,U}$ and,

[17] For detailed discussions regarding finite dimensional or infinite dimensional sieve spaces, see Chen (2005) and Shen (1997, 1998).

[18] Note that with some abuse of notation, when we consider $\widehat{\mathcal{H}}_{l,\delta_n}$ as a sequence of sets indexed by l, we treat $\widehat{h}(\cdot,\theta)$ is fixed. In other words, the degree of approximation of the sieve space for the estimation stage does not have to agree with that of the sieve space for the inference, which we can let arbitrary large regardless of the sample size.

analogously, $\widetilde{c}_{n,L}$ is an upper bound on c_0 for $(j,m) \in \widehat{\mathcal{B}}_{n,L}$ at $\alpha = \widehat{\alpha}_{n,L}$. Thus, $\widetilde{c}_{n,U}$ and $\widetilde{c}_{n,L}$ are given as random real functions on $\mathcal{A}$:

$$\begin{aligned} \widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha) &= \widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\alpha) + \widehat{w}_n(j,\widehat{\gamma}_{n,j,m},\alpha)\lambda^*_{n,U}(j,m,a)/\sqrt{n} \text{ and} \\ \widetilde{c}_{n,L}(j,\widehat{\gamma}_{n,j,m},\alpha) &= \widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\alpha) + \widehat{w}_n(j,\widehat{\gamma}_{n,j,m},\alpha)\lambda^*_{n,L}(j,m,a)/\sqrt{n}, \end{aligned} \tag{22}$$

where $\widehat{w}_n(j,\gamma,\alpha)$ is a positive weight function[19] and $\lambda^*_{n,U}(j,m,a)\&\lambda^*_{n,L}(j,m,a)$ are non-negative critical values that are constructed by the bootstrap procedure described in the following section.

Now we compare the three alternative CI's. Alternative CI1 is most computationally demanding since we construct the bounds over the whole infinite dimensional parameter space. However, we note that this is not too demanding compared to the estimation stage since the construction of the bounds do not involve the optimization. Moreover, we do not require some slackness variable such as δ_n. For Alternative CI2, we include the slackness variable δ_n in constructing $\widetilde{\beta}_{n,L}$ and $\widetilde{\beta}_{n,U}$ so that while reducing the functional space for such construction from $\Theta \times \mathcal{H}$ to $\Theta \times \widehat{\mathcal{H}}_{\delta_n}$, we make sure $\Theta_+ \times \mathcal{H}_+$ is included in $\Theta \times \widehat{\mathcal{H}}_{\delta_n}$ with probability approaching to one. This is critical in justifying the proposed CI asymptotically. In practice, we can let δ_n be some fixed number since fixed δ_n does not affect the asymptotics for the CI. We still have a valid CI with a fixed δ_n. However, the choice of δ_n will affect the cost of computation and thus we want to let δ_n be small as long as the sample size is relatively large. Alternative CI3 requires the least computation among three alternatives since the CI is constructed over the finite dimensional sieve space but we need to admit the possibility that the coverage probability is smaller than $1-a$ with a finite l. However, in practice, we can let l be arbitrary large noting the sieve space for the construction of the bounds can be larger than that for the estimation stage and thus the size of data does not restrict the smoothness of the sieve space for the inference stage. In consequence, we can make the smallest value of coverage probability arbitrary close to $1-a$.

Now we consider how to obtain the critical values $\lambda^*_{n,U}(j,m,a)$ and $\lambda^*_{n,L}(j,m,a)$ using the standard nonparametric bootstrap.

5.1 Bootstrap Critical Values

Here we briefly review the bootstrap procedure to obtain the critical values following ABJ. Most of their discussions hold by replacing their θ and $\widehat{\theta}$ with the infinite dimensional parameter α and $\widehat{\alpha}$. The purpose of our paper is to provide some conditions under which the bootstrap critical values can be justified asymptotically for the semiparametric case.

Let $\{(Y_i^*, X_i^*) : i = 1,\ldots,n\}$ denote a standard nonparametric bootstrap sample conditional on the original sample $\{(Y_i, X_i) : i = 1,\ldots,n\}$. First, we obtain the bootstrap conditional probabilities implied by the model using the bootstrap sample such that $P^*(y^{(j)}|X_i^*,\theta,h(\cdot_i^*)) = P(y^{(j)}|X_i^*,\theta,h(\cdot_i^*))$. Similarly, we define $\widehat{c}^*_n(j,\gamma,\alpha)$, $\widehat{\gamma}^*_{n,j,m}$, $\widehat{\Gamma}^*_n$, and $\widehat{w}^*_n(j,\gamma,\theta)$ using the bootstrap sample as we define $\widehat{c}_n(j,\gamma,\alpha)$, $\widehat{\gamma}_{n,j,m}$, $\widehat{\Gamma}_n$, and $\widehat{w}_n(j,\gamma,\alpha)$, respectively. Define

$$\begin{aligned} D^*_{n,U}(j,m) &= \sqrt{n}\left(\left(\widehat{c}^*_n(j,\widehat{\gamma}^*_{n,j,m},\widehat{\alpha}_{n,U}) - \widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right)/\, w^*_n\left(j,\widehat{\gamma}^*_{n,j,m},\widehat{\alpha}_{n,U}\right)\right) \text{ and} \\ D^*_{n,L}(j,m) &= \sqrt{n}\left(\left(\widehat{c}^*_n(j,\widehat{\gamma}^*_{n,j,m},\widehat{\alpha}_{n,L}) - \widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,L})\right)/\, w^*_n\left(j,\widehat{\gamma}^*_{n,j,m},\widehat{\alpha}_{n,L}\right)\right). \end{aligned}$$

Note that when we construct $D^*_{n,U}(j,m)$ and $D^*_{n,L}(j,m)$, in the arguments of $\widehat{c}^*_n(\cdot)$, we use $\widehat{\alpha}_{n,U}$ and $\widehat{\alpha}_{n,L}$ not $\widehat{\alpha}^*_{n,U}$ and $\widehat{\alpha}^*_{n,L}$, respectively. This ensures that $E^*\left[\widehat{c}^*_n(j,\widehat{\gamma}^*_{n,j,m},\widehat{\alpha}_{n,U}) - \widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right] = 0$ and $E^*\left[\widehat{c}^*_n(j,\widehat{\gamma}^*_{n,j,m},\widehat{\alpha}_{n,L}) - \widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,L})\right] = 0$ where E^* is the expectation operator taken conditional on the original sample. We denote $P^*(\cdot)$ to be the probability with respect to the bootstrap sample conditional

[19] The weight function is used to make $\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\alpha)/\widehat{w}_n(j,\widehat{\gamma}_{n,j,m},\alpha)$ have comparable distributions across different (j,m). Examples of such weight functions can be found in ABJ.

on the original sample. Now we obtain the critical values $\lambda^*_{n,U}(j,m,a)$ for $(j,m) \in \widehat{\mathcal{B}}_{n,U}$ and $\lambda^*_{n,L}(j,m,a)$ for $(j,m) \in \widehat{\mathcal{B}}_{n,L}$ as non-negative constants satisfying the condition

$$P^* \left(\begin{array}{l} D^*_{n,U}(j,m) + \lambda^*_{n,U}(j,m,a) \geq 0 \text{ for } (j,m) \in \widehat{\mathcal{B}}_{n,U}) \text{ and} \\ D^*_{n,U}(j,m) + \lambda^*_{n,L}(j,m,a) \geq 0 \text{ for } (j,m) \in \widehat{\mathcal{B}}_{n,L}) \end{array} \right) = 1 - a \tag{23}$$

and the same condition with U and L interchanged.[20] We summarize the procedure to construct the CI's.

1 Obtain the following objects in the order that they are presented:
$\widehat{\Gamma}_n$, $\widehat{c}_n(j,\widehat{\gamma}_{n,k,m},\alpha)$, $\widehat{\mathcal{A}}_n$, $\widehat{\beta}_{n,U}, \widehat{\beta}_{n,L}$, $\widehat{\alpha}_{n,U}, \widehat{\alpha}_{n,L}$, $\widehat{\mathcal{B}}_{n,U}, \widehat{\mathcal{B}}_{n,L}$, $\widehat{w}_n(j,\widehat{\gamma}_{n,k,m},\widehat{\alpha}_{n,U})$, $\widehat{w}_n(j,\widehat{\gamma}_{n,k,m},\widehat{\alpha}_{n,L})$.

2 Obtain the bootstrap critical values as described in this section:
$\lambda^*_{n,U}(j,m,a)$ for $(j,m) \in \widehat{\mathcal{B}}_{n,U}$ and $\lambda^*_{n,L}(j,m,a)$ for $(j,m) \in \widehat{\mathcal{B}}_{n,L}$

3 Construct the confidence intervals defined in (17):
Obtain $\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,k,m},\widehat{\alpha}_{n,U})$, $\widetilde{c}_{n,L}(j,\widehat{\gamma}_{n,k,m},\widehat{\alpha}_{n,L})$, $\widetilde{\beta}_{n,U}$, and $\widetilde{\beta}_{n,L}$ from (22) and (19), (20), or (21), respectively.

Under some higher level assumptions presented in the Appendix, the following theorem justifies the confidence intervals suggested in (17) asymptotically for all the three alternatives.

Theorem 5.1 *Suppose Assumptions B.1-B.6 in the Appendix are satisfied. Further suppose Assumptions 4.1, 4.2, 4.3, and 4.7 hold. Then, Alternative CI1 satisfies (18) with* $\widetilde{\beta}_{n,U} = \widetilde{\beta}^{(1)}_{n,U}$ *and* $\widetilde{\beta}_{n,L} = \widetilde{\beta}^{(1)}_{n,L}$. *Further suppose* $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \underset{p}{\to} 0$. *Then, (18) with* $\widetilde{\beta}_{n,U} = \widetilde{\beta}^{(2)}_{n,U}$ *and* $\widetilde{\beta}_{n,L} = \widetilde{\beta}^{(2)}_{n,L}$ *holds for Alternative CI2. Further suppose* $\widetilde{\beta}^{(3)}_{l,n,\cdot}$ *has a well-defined cdf whose derivative is uniformly bounded over its support. Then, we have* $\liminf_{l,n\to\infty, l\geq n} P(\beta_0 \subset CI_n(1-a)) = \liminf_{l,n\to\infty, l\geq n} P(\widetilde{\beta}^{(3)}_{l,n,L} \leq \beta_0 \leq \widetilde{\beta}^{(3)}_{l,n,U}) \geq 1-a.$

6 Concluding Remarks

This paper considers estimation and inference of parameters in discrete games with multiple equilibria, without using an equilibrium selection rule, while the game model can contain infinite dimensional parameters. In particular, we adopt a set inference approach popularized recently. Noting the literature only allows for finite dimensional parameters in the model even though infinite dimensional parameter is naturally included in the model or misspecification of a fully parametric model is concerned, this paper extends a current literature to a set inference with infinite dimensional parameters where a consistent profiled estimator of infinite dimensional parameters is available. A consistent set estimator and confidence intervals are provided. Examples of signaling games with discrete types where the type distribution is nonparametrically specified and entry-exit games with partially linear payoffs functions are considered.

In this paper, we note that achieving a consistent point estimation often requires some information reduction (For example, redefinition of outcome spaces). Due to this less use of information than available in the model, we may end up a point estimator with larger variances and have wider confidence intervals than those of the set estimator using the full information in the model. This finding justifies the use of the set inference even though we can achieve a consistent point estimation in some cases. It is also an interesting future research to compare these two alternatives: CI from the point estimation with the usage of less information vs. CI from the set estimation with the usage of the full information.

[20] Though the requirement of (23) for $\lambda^*_{n,L}(\cdot,\cdot,\cdot)$ and $\lambda^*_{n,U}(\cdot,\cdot,\cdot)$ is enough to justify the CI asymptotically, it does not uniquely determine these values. Also in principle, these bootstrapped critical values can be obtained analytically but in practice, they need to be simulated. See some related issues in ABJ. We omit these discussions since our focus here is to provide an asymptotic justification for the semiparameric version of the CI proposed in ABJ.

Appendix

A Consistency Proofs

For any real functional c on $\mathcal{I}_J \times \Gamma_{all} \times \Theta \times \mathcal{H}_{\delta_n}$ and any collection of $\sum_{j=1}^{J} \mathcal{M}_j$ subsets of Γ of $\mathbb{S}(X)$, define $\Theta(c, \Gamma, h) = \{\theta \in \Theta : c(j, \gamma_{j,m}, \theta, h(\cdot, \theta)) \geq 0, \forall (j,m) \in \mathcal{I}_{J,\mathcal{M}}\}$ and note that $\Theta(c_0, \Gamma_0, h_0) = \Theta_+$ and that if $\Theta\left(\widehat{c}_n, \widehat{\Gamma}_n, \widehat{h}\right)$ is non-empty, we have $\Theta\left(\widehat{c}_n, \widehat{\Gamma}_n, \widehat{h}\right) = \widehat{\Theta}_n$, which implies that $\Theta\left(\widehat{c}_n, \widehat{\Gamma}_n, \widehat{h}\right) \subset \widehat{\Theta}_n$. To prove Theorem 4.1, we need the following lemma which extends Lemma 4 in ABJ to the semiparametric case.

Lemma A.1 *Under Assumptions 4.4 and 4.5 (ii), $\rho(\Theta(c_0, \Gamma_0, h_0) | \Theta(c, \Gamma, h)) \to 0$ as $c \to c_0$, $\Gamma \to \Gamma_0$, $h \to h_0$ under $\|\cdot\|_{U_n}$, $\|\cdot\|$, and $\|\cdot\|_{\mathcal{H}}$.*

Proof. For any $\theta \in \Theta$, we have

$$\begin{aligned} &\limsup_{(c,\Gamma,h)\to(c_0,\Gamma_0,h_0)} \left|\min_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} c(j, \gamma_{j,m}, \theta, h) - \min_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} c_0(j, \gamma_{j,m}, \theta, h)\right| \\ &\leq \limsup_{c\to c_0} \sup_{h\in\mathcal{H}_{\delta_n}} \sup_{j\in\mathcal{I}_J, \gamma\in\Gamma_{all}} |c(j, \gamma, \theta, h) - c_0(j, \gamma, \theta, h)| = 0, \end{aligned}$$

because $c \to c_0$ with respect to the uniform metric over $\mathcal{I}_J \times \Gamma_{all} \times \Theta \times \mathcal{H}_{\delta_n}$. It follows that for any $\theta \in \Theta$,

$$\liminf_{(c,\Gamma,h)\to(c_0,\Gamma_0,h_0)} \min_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} c(j, \gamma_{j,m}, \theta, h) = \liminf_{\Gamma\to\Gamma_0, h\to h_0} \min_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} c_0(j, \gamma_{j,m}, \theta, h). \tag{24}$$

Next, consider that for any $\alpha \in \Theta \times \mathcal{H}_{\delta_n}$, we have

$$\begin{aligned} &\limsup_{\Gamma\to\Gamma_0} \left|\min_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} c_0(j, \gamma_{j,m}, \alpha) - \min_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} c_0(j, \gamma_{0,j,m}, \alpha)\right| \\ &\leq \limsup_{\Gamma\to\Gamma_0} \max_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} 2\int \left|q_{\gamma_{j,m}}(x) - q_{\gamma_{0,j,m}}(x)\right| dF_X(x) = 0, \end{aligned}$$

where the inequality holds by the definition of c_0 in (3) and the equality holds by the definition of $\Gamma \to \Gamma_0$ in (12) and the Cauchy-Schwarz inequality. Also note for any $\Gamma \subset \Gamma_{all}$ and $\theta \in \Theta$, we have

$$\begin{aligned} &\limsup_{h\to h_0} \left|\min_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} c_0(j, \gamma_{j,m}, \theta, h(\cdot, \theta)) - \min_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} c_0(j, \gamma_{j,m}, \theta, h(\cdot, \theta))\right| \\ &\leq \limsup_{h\to h_0} \left(\sup_{x\in\mathbb{S}(X)} \max_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} \left|q_{\gamma_{j,m}}(x)\right|\right) \|h - h_0\|_{\mathcal{H}} = 0, \end{aligned}$$

where the inequality holds by the construction of $c_0(\cdot)$ and Assumption 4.4 (i) and the equality holds by the definition of the metric $\|\cdot\|_{\mathcal{H}}$ and Assumption 4.4 (ii). It follows that for any $\theta \in$int(Θ_+),

$$\liminf_{\Gamma\to\Gamma_0, h\to h_0} \min_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} c_0(j, \gamma_{j,m}, \theta, h) = \min_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} c_0(j, \gamma_{0,j,m}, \theta, h_0) > 0, \tag{25}$$

where the last result holds by Assumption 4.5 (ii). From (24) and (25), we conclude that for any $\theta \in$int(Θ_+), it also holds that $\theta \in \Theta(c, \Gamma, h)$ for (c, Γ, h) sufficiently close to (c_0, Γ_0, h_0). Now suppose $\rho(\Theta(c_0, \Gamma_0, h_0) | \Theta(c, \Gamma, h)) \nrightarrow 0$ as $(c, \Gamma, h) \to (c_0, \Gamma_0, h_0)$. Then, by definition of $\rho(\Theta(c_0, \Gamma_0, h_0) | \Theta(c, \Gamma, h)) = \sup_{\theta\in\Theta(c_0,\Gamma_0,h_0)} \rho(\theta | \Theta(c, \Gamma, h))$, there exists (i) a constant $\varepsilon > 0$, (ii) a sequence of functions on $\mathcal{I}_J \times \Gamma_{all} \times \Theta \times \mathcal{H}_{\delta_n}$, $\{c_j : j \geq 1\}$, and a sequence of collections of $\sum_{j=1}^{J} \mathcal{M}_j$ sets in Γ_{all}, $\{\Gamma_j : j \geq 1\}$, s.t. $(c_j, \Gamma_j, h_j) \to (c_0, \Gamma_0, h_0)$, and (iii) a sequence of parameters $\left\{\theta_{c_j} \in \Theta(c_0, \Gamma_0, h_0) : j \geq 1\right\}$ s.t. $\rho\left(\theta_{c_j} | \Theta(c_j, \Gamma_j, h_j)\right) \geq \varepsilon$ for all $j \geq 1$. The sequence $\left\{\theta_{c_j} \in \Theta(c_0, \Gamma_0, h_0) : j \geq 1\right\}$ has a subsequence, say $\left\{\theta_{c_{j_l}} : l \geq 1\right\}$, that converges to a point $\Theta(c_0, \Gamma_0, h_0)$ because $\Theta(c_0, \Gamma_0, h_0)$ is compact (This is because Θ is compact and $\Theta(c_0, \Gamma_0, h_0)$ is defined from the non-strict inequality.). That is, $\left\|\theta_{c_{j_l}} - \theta_*\right\|_E \to 0$ as $l \to \infty$ for some $\theta_* \in \Theta(c_0, \Gamma_0, h_0)$. For all l sufficiently large satisfying $\left\|\theta_{c_{j_l}} - \theta_*\right\|_E < \varepsilon/2$, we have

$\left|\rho\left(\theta_{c_{j_l}}|\Theta\left(c_{j_l},\Gamma_{j_l},h_{j_l}\right)\right)-\rho\left(\theta_*|\Theta\left(c_{j_l},\Gamma_{j_l},h_{j_l}\right)\right)\right| \leq \left\|\theta_{c_{j_l}}-\theta_*\right\|_E < \varepsilon/2$ by the triangle inequality. Thus, for all l sufficiently large,

$$\rho\left(\theta_*|\Theta\left(c_{j_l},\Gamma_{j_l}\right)\right) \geq \rho\left(\theta_{c_{j_l}}|\Theta\left(c_{j_l},\Gamma_{j_l}\right)\right)-\varepsilon/2 \geq \varepsilon/2. \tag{26}$$

If $\theta_* \in \text{int}(\Theta_+)$, (26) contradicts to the fact that for any $\theta \in \text{int}(\Theta_+)$, $\theta \in \Theta\left(c,\Gamma,h\right)$ for (c,Γ,h) sufficiently close to (c_0,Γ_0,h_0). If $\theta_* \notin \text{int}(\Theta_+)$, then by Assumption 4.5 (ii)(a), there exists a $\theta_{\text{int}} \in \text{int}(\Theta_+)$ such that $\|\theta_*-\theta_{\text{int}}\|_E < \varepsilon/4$. Applying the triangle inequality to (26), we obtain $\rho\left(\theta_{\text{int}}|\Theta\left(c_{j_l},\Gamma_{j_l},h_{j_l}\right)\right) \geq \varepsilon/4$ for all l sufficiently large. This is also a contradiction. We conclude that $\rho\left(\Theta\left(c_0,\Gamma_0,h_0\right)|\Theta\left(c,\Gamma,h\right)\right) \to 0$ as $(c,\Gamma,h) \to (c_0,\Gamma_0,h_0)$. ■

A.1 Proof of Theorem 4.1

We first prove part (i) by extending a consistency result of a class of extremum estimator. Under Assumption 4.4, $Q(\theta,h)$ defined in (7) is continuous (with respect to the metric $\|\cdot\|_s$). Note that this holds even though $Q(\theta,h)$ contains an indicator function because $|b(\theta,h)|\cdot\mathbf{1}[b(\theta,h)\leq 0]$ is continuous as long as $b(\theta,h)$ is continuous (with respect to the metric $\|\cdot\|_s$), which follows from the fact that $|b(\theta_1,h_1)|\cdot\mathbf{1}[b(\theta_1,h_1)\leq 0]-|b(\theta_2,h_2)|\cdot\mathbf{1}[b(\theta_2,h_2)\leq 0] \leq ||b(\theta_1,h_1)|-|b(\theta_2,h_2)||$ for all (θ_1,h_1), (θ_2,h_2). Because $Q(\theta,h)$ is continuous and $\Theta\times\mathcal{H}$ is compact under $\|\cdot\|_s$, $Q(\cdot,\cdot)$ attains its minimum value zero at points in the set $\mathcal{A}_+$ by definition in (8). Now we claim that for all $\varepsilon>0$, there exists a $\delta>0$ such that

$$\inf_{\theta\notin\mathcal{E}\left(\Theta_+,\varepsilon\right),\theta\in\Theta,\|h-h_0\|_{\mathcal{H}}\leq\delta_n,h\in\mathcal{H}} Q(\theta,h) \geq \delta > 0 \tag{27}$$

where $\mathcal{E}\left(\Theta_+,\varepsilon\right)=\{\theta\in\Theta:\rho(\theta|\Theta_+)<\varepsilon\}$. Suppose not. Then, for some $\varepsilon>0$ and h s.t. $\|h-h_0\|_{\mathcal{H}}\leq\delta_n$, there is a sequence $\{\theta_l\in\Theta/\mathcal{E}\left(\Theta_+,\varepsilon\right):l\geq 1\}$ for which $\lim_{l\to\infty} Q(\theta_l,h(\cdot,\theta_l))=0$. Because Θ is compact and $\mathcal{E}\left(\Theta_+,\varepsilon\right)$ is open, the set $\Theta/\mathcal{E}\left(\Theta_+,\varepsilon\right)$ is compact. Hence, $\{\theta_l:l\geq 1\}$ has a convergent subsequence, say $\left\{\theta_{l_j}:j\geq 1\right\}$, that converges to a point in $\Theta/\mathcal{E}\left(\Theta_+,\varepsilon\right)$, say θ_∞. Continuity of $Q(\cdot,\cdot)$ and $h(\cdot,\theta)\in\mathcal{H}_{\delta_n}$ in θ imply that $Q\left(\theta_\infty,h(\cdot,\theta_\infty)\right)=\lim_{j\to\infty} Q\left(\theta_{l_j},h(\cdot,\theta_{l_j})\right)=0$. This implies that $\theta_\infty\in\Theta_+$, which is a contradiction. This proves (27). From $\mathcal{H}_n\subseteq\mathcal{H}_{n+1}\subseteq\mathcal{H}$ by Assumption 4.3 (ii) (note also $\mathcal{H}_n$ is compact) and the fact that $Q(\theta,h)$ and $h(\cdot,\theta)$ are continuous in (θ,h) and θ, respectively, we note that there is an $N\geq 1$ such that $\inf_{\theta\notin\mathcal{E}\left(\Theta_+,\varepsilon\right),\theta\in\Theta,h\in\mathcal{H}_{n,\delta_n}} Q(\alpha) \geq \inf_{\theta\notin\mathcal{E}\left(\Theta_+,\varepsilon\right),\theta\in\Theta,h\in\mathcal{H}_{\delta_n}} Q(\alpha)$ for all $n\geq N$, recalling $\mathcal{H}_{\delta_n}\equiv\{h\in\mathcal{H}:\|h(\cdot,\theta)-h_0(\cdot,\theta)\|_{\mathcal{H}}\leq\delta_n,\ \theta\in\Theta\}$ and $\mathcal{H}_{n,\delta_n}\equiv\{h\in\mathcal{H}_n:\|h(\cdot,\theta)-h_0(\cdot,\theta)\|_{\mathcal{H}}\leq\delta_n,\ \theta\in\Theta\}$ with $\delta_n=o(1)$. From this result and (27), it follows that

$$\inf_{\theta\notin\mathcal{E}\left(\Theta_+,\varepsilon\right),\theta\in\Theta,h\in\mathcal{H}_{n,\delta_n}} Q(\alpha) \geq \delta > 0 \text{ for all } n\geq N. \tag{28}$$

This is a version of the identification condition. Now we derive the uniform convergence of $Q_n(\theta,h)$ defined in (6) to $Q(\theta,h)$ uniformly over $\Theta\times\mathcal{H}_{\delta_n}$. Consider

$$\begin{aligned}
\sup_{\theta\in\Theta,h\in\mathcal{H}_{\delta_n}} |Q_n(\theta,h)-Q(\theta,h)| &\leq \sup_{\theta\in\Theta,h\in\mathcal{H}_{\delta_n}} \left\{\max_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} \left|\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\theta,h)-c_0(j,\gamma_{0,j,m},\theta,h)\right|\right\} \\
&\leq \sup_{\theta\in\Theta,h\in\mathcal{H}_{\delta_n}} \left\{\max_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} \left|\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\theta,h)-c_0(j,\widehat{\gamma}_{n,j,m},\theta,h)\right|\right\} \\
&\quad + \sup_{\theta\in\Theta,h\in\mathcal{H}_{\delta_n}} \left\{\max_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} \left|c_0(j,\widehat{\gamma}_{n,j,m},\theta,h)-c_0(j,\gamma_{0,j,m},\theta,h)\right|\right\} \\
&\leq \sup_{(j,\gamma,\theta,h)\in\mathcal{I}_J\times\Gamma_{all}\times\Theta\times\mathcal{H}_{\delta_n}} \left|\widehat{c}_n(j,\gamma,\theta,h)-c_0(j,\gamma,\theta,h)\right| + 2\max_{(j,m)\in\mathcal{I}_{J,\mathcal{M}}} \int\left|q_{\widehat{\gamma}_{n,j,m}}(x)-q_{\gamma_{0,j,m}}(x)\right| dF_X(x) \underset{p}{\to} 0
\end{aligned} \tag{29}$$

where the first inequality holds by the definitions of $Q_n(\theta,h)$ and $Q(\theta,h)$, the second inequality is from the triangle inequality, the third inequality is obtained using the definition of $c_0(j,\gamma,\theta,h)$ in (3), and last result holds by Assumptions 4.6 (i) and 4.7 using the definitions of the metric $\|\cdot\|_{U_n}$ and $\|\cdot\|$. Now we are ready to prove Theorem 4.1 (i). Note that the set Θ_+ is not empty by Assumption 4.2 and $\widehat{\Theta}_n$ is not empty by construction. Let $\alpha_{n+}=(\theta_+,h_{n+})$ denote some element of $\Theta_+\times\left\{h\in\mathcal{H}_n:\|h(\cdot,\theta)-h_0(\cdot,\theta)\|_{\mathcal{H}}\leq\delta_n,\theta\in\Theta_+\right\}$. Then, there exist $\alpha_+\in\mathcal{A}_+$ such that $\|\alpha_{n+}-\alpha_+\|_s=o(1)$ and thus $Q(\alpha_{n+})-\delta/2\leq Q(\alpha_+)$ for $n\geq\exists N$ by the continuity of $Q(\cdot,\cdot)$. It follows that

$$-Q(\alpha_{n+}) \geq -\delta/2 \text{ for } n\geq\exists N \tag{30}$$

since $Q(\alpha_+) = 0$ for every $\alpha_+ \in \mathcal{A}_+$. (28) and the fact that Θ_+ and $\widehat{\Theta}_n$ are not empty imply that for all $\varepsilon > 0$ there exists a $\delta > 0$ such that

$$\begin{aligned}
&P\left(\rho\left(\widehat{\Theta}_n|\Theta_+\right) > \varepsilon\right) = P\left(\widehat{\Theta}_n \cap (\Theta / \mathcal{E}(\Theta_+, \varepsilon)) \neq \varnothing\right) \\
&\leq P\left(\sup_{\theta \in \widehat{\Theta}_n, h \in \mathcal{H}_{n,\delta_n}} Q(\alpha) \geq \delta\right) = P\left(\sup_{\theta \in \widehat{\Theta}_n, h \in \mathcal{H}_{n,\delta_n}} (Q(\alpha) - Q_n(\alpha) + Q_n(\alpha)) \geq \delta\right) \\
&\leq P\left(\sup_{\theta \in \widehat{\Theta}_n, h \in \mathcal{H}_{n,\delta_n}} (Q(\alpha) - Q_n(\alpha) + Q_n(\alpha_{n+}) - Q(\alpha_{n+}) + o(1)) \geq \delta/2\right) \\
&\leq P\left(2 \sup_{\theta \in \Theta, h \in \mathcal{H}_{\delta_n}} |Q_n(\alpha) - Q(\alpha)| \geq \delta/2\right) \to 0
\end{aligned}$$

where the first inequality holds by (28), the second inequality holds because (i) $Q_n(\alpha) \leq Q_n(\alpha_{n+})$ for each $\alpha \in \widehat{\mathcal{A}}_n$ since each $\alpha \in \widehat{\mathcal{A}}_n$ minimizes $Q_n(\alpha)$ over $\Theta \times \mathcal{H}_{\delta_n}$ and (ii) $-Q(\alpha_{n+}) + o(1) \geq -\delta/2$ as noted in (30). The last result comes from (29) and $\mathcal{H}_n \subseteq \mathcal{H}_{n+1} \subseteq \mathcal{H}$ for all n. This completes the proof of part (i).

Now we turn to part (ii). Suppose Assumption 4.5 (ii) holds. Then, by Lemma A.1, given $\varepsilon > 0$, there exists $\delta > 0$ such that $\|c - c_0\|_{U_n} < \delta$, $\|\Gamma - \Gamma_0\| < \delta$, and $\|h - h_0\|_{\mathcal{H}} < \delta$ imply that $\rho(\Theta(c_0, \Gamma_0, h_0)|\Theta(c, \Gamma, h)) < \varepsilon$ as $n \to \infty$. It follows that

$$P\left(\rho\left(\Theta(c_0, \Gamma_0, h_0)|\Theta(\widehat{c}_n, \widehat{\Gamma}_n, \widehat{h})\right) < \varepsilon\right) \geq P\left(\|\widehat{c}_n - c_0\|_{U_n} < \delta, \left\|\widehat{\Gamma}_n - \Gamma_0\right\| < \delta, \left\|\widehat{h} - h_0\right\|_{\mathcal{H}}\right) \to 1, \tag{31}$$

where the convergence holds by Assumption 4.3 (iii), 4.6, and 4.7. From $\Theta\left(\widehat{c}_n, \widehat{\Gamma}_n, \widehat{h}\right) \subset \widehat{\Theta}_n$, it follows that

$$P\left(\rho\left(\Theta(c_0, \Gamma_0, h_0)|\widehat{\Theta}_n\right) < \varepsilon\right) \geq P\left(\rho\left(\Theta(c_0, \Gamma_0, h_0)|\Theta_n(\widehat{c}_n, \widehat{\Gamma}_n, \widehat{h})\right) < \varepsilon\right). \tag{32}$$

Combining (31) and (32), we establish the part (ii) of Theorem 4.1 under Assumption 4.5 (ii).

Next, suppose Assumption 4.5 (i) holds. Then, $\rho\left(\Theta|\widehat{\Theta}_n\right) = \rho\left(\{\theta_0\}|\widehat{\Theta}_n\right) \leq \rho\left(\widehat{\Theta}_n|\{\theta_0\}\right) \underset{p}{\to} 0$ where the inequality holds since (i) the distance from a point to a non-empty set is less than or equal to the distance from the set to the point by definition of $\rho(\cdot|\cdot)$ and (ii) the set $\widehat{\Theta}_n$ is not empty. The convergence result holds by the proof of part (i).

Now we show that Assumption 4.3 (iv) and $d(\widehat{\Theta}_n, \Theta_+) \to 0$ imply $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \to 0$. Note for any $\alpha = (\theta, h(\cdot, \theta)) \in \widehat{\mathcal{A}}_n$, we can find $\exists \alpha' = (\theta', h_0(\cdot, \theta')) \in \mathcal{A}_+$ s.t. $\|\alpha - \alpha'\|_s = \|\theta - \theta'\|_E + \|h(\cdot, \theta) - h_0(\cdot, \theta')\|_{\mathcal{H}} \to 0$ since $\|\theta - \theta'\|_E \to 0$ by $d(\widehat{\Theta}_n, \Theta_+) \to 0$ and since $\|h(\cdot, \theta) - h_0(\cdot, \theta')\|_{\mathcal{H}} \leq \|h(\cdot, \theta) - h_0(\cdot, \theta)\|_{\mathcal{H}} + \|h_0(\cdot, \theta') - h_0(\cdot, \theta)\|_{\mathcal{H}} \to 0$ because $h(\cdot, \theta) \in \widehat{\mathcal{H}}_n \subset \mathcal{H}_{\delta_n}$ by definition and because $h_0(\cdot, \theta)$ is continuous in θ and $d(\widehat{\Theta}_n, \Theta_+) \to 0$. It follows that $\rho(\widehat{\mathcal{A}}_n|\mathcal{A}_+) \to 0$. Similarly for any $\alpha = (\theta, h_0(\cdot, \theta)) \in \mathcal{A}_+$, we can find $\exists \alpha' = (\theta', h(\cdot, \theta')) \in \widehat{\mathcal{A}}_n$ such that $\|\alpha - \alpha'\|_s \to 0$, which implies $\rho(\mathcal{A}_+|\widehat{\mathcal{A}}_n) \to 0$. This completes the proof of Theorem 4.1.

A.2 Consistency of $\widehat{\beta}_{n,U}$ and $\widehat{\beta}_{n,L}$ (Proof of Corollary 4.1)

The following proof is essentially the same with that of Theorem 2 in ABJ and can be omitted.

For any two sets of real numbers B_1 and B_2, let $b_j^* = \sup\{b \in B_j\}$ for $j = 1, 2$ and note that[21] $|b_1^* - b_2^*| \leq d(B_1, B_2)$. Now define $\widehat{B}_n = \left\{\beta_n(\alpha) : \alpha \in \widehat{\mathcal{A}}_n\right\}$ and $B_{n,+} = \{\beta_n(\alpha) : \alpha \in \mathcal{A}_+\}$. Then, from the result above, it follows that

$$\left|\widehat{\beta}_{n,U} - \beta_{n,U}\right| \leq d\left(\widehat{B}_n, B_{n,+}\right) \tag{33}$$

by definitions of $\widehat{\beta}_{n,U}$ and $\beta_{n,U}$ given in (13) and (14), respectively. Now we note that

$$P(d(\widehat{B}_n, B_{n,+}) > \varepsilon) \leq P(\rho(\widehat{B}_n|B_{n,+}) > \varepsilon) + P(\rho(B_{n,+}|\widehat{B}_n) > \varepsilon) \tag{34}$$

[21] To prove this, suppose $b_1^* > b_2^*$. Then, $|b_1^* - b_2^*| = \rho(b_1^*|B_2) \leq \sup_{b \in B_1} \rho(b|B_2) = \rho(B_1|B_2) \leq d(B_1, B_2)$. Analogously, we can show this is true for $b_1^* < b_2^*$.

from the definition of $d(\cdot,\cdot)$. Let $p_\delta = P(d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) > \delta)$ and consider

$$\begin{aligned} &P(\rho(\widehat{B}_n|B_{n,+}) > \varepsilon) \leq P(\rho(\widehat{B}_n|B_{n,+}) > \varepsilon, d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \leq \delta) + p_\delta \\ &= P(\sup_{\alpha \in \widehat{\mathcal{A}}_n} \inf\{|\beta_n(\alpha) - \beta_n(\alpha_+)| : \alpha_+ \in \mathcal{A}_+\} > \varepsilon, d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \leq \delta) + p_\delta \end{aligned} \tag{35}$$

where the equality holds by definitions of $\rho(\cdot|\cdot)$, $\widehat{B}_n$, and $B_{n,+}$. Now consider that if $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \leq \delta$, then $\rho(\widehat{\mathcal{A}}_n|\mathcal{A}_+) \leq \delta$ and for any $\alpha \in \widehat{\mathcal{A}}_n$, there exists $\alpha_{++} \in \mathcal{A}_+$ such that $\|\alpha - \alpha_{++}\|_s \leq \delta$. It follows that

$$\begin{aligned} &\sup_{\alpha \in \widehat{\mathcal{A}}_n} \inf\{|\beta_n(\alpha) - \beta_n(\alpha_+)| : \alpha_+ \in \mathcal{A}_+\} \\ &\leq \sup_{\alpha \in \widehat{\mathcal{A}}_n, \|\alpha - \alpha_{++}\|_s \leq \delta} |\beta_n(\alpha) - \beta_n(\alpha_{++})| \leq \sup_{\|\alpha_1 - \alpha_2\|_s \leq \delta} |\beta_n(\alpha_1) - \beta_n(\alpha_2)| \end{aligned} \tag{36}$$

where the second inequality holds since $\|\alpha - \alpha_{++}\|_s \leq \delta$ for any $\alpha \in \widehat{\mathcal{A}}_n$. From (35), (36), and Assumption 4.8, it follows that for sufficient large n, $P(\sup_{\alpha \in \widehat{\mathcal{A}}_n} \inf\{|\beta_n(\alpha) - \beta_n(\alpha_+)| : \alpha_+ \in \mathcal{A}_+\} > \varepsilon, d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \leq \delta) \leq \varepsilon$ and thus

$$P(\rho(\widehat{B}_n|B_{n,+}) > \varepsilon) \leq \varepsilon + P(d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) > \delta). \tag{37}$$

An analogous argument provides the same result as (37) but replacing $\rho(\widehat{B}_n|B_{n,+})$ & $\rho(\widehat{\mathcal{A}}_n|\mathcal{A}_+)$ with $\rho(B_{n,+}|\widehat{B}_n)$ & $\rho(\mathcal{A}_+|\widehat{\mathcal{A}}_n)$, respectively. To be precise, $P(\rho(B_{n,+}|\widehat{B}_n) > \varepsilon) \leq \varepsilon + P(d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) > \delta)$. Combining this with (33), (34), and (37), we obtain $P(|\widehat{\beta}_{n,U} - \beta_{n,U}| > \varepsilon) \leq P(d(\widehat{B}_n, B_{n,+}) > \varepsilon) \leq 2\varepsilon + 2P(d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) > \delta)$. This proves Corollary 4.1 since $\varepsilon > 0$ is arbitrary and $P(d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) > \delta) \to 0$ by Theorem 4.1.

A.3 Estimation of the Type Distribution: Proof of Proposition 4.1

We derive the consistency for the set estimator of the type distribution parameter. We let $p_{u,n}(\alpha) = \frac{1}{n}\sum_{i=1}^{n} L(h(z_i))$ with $\alpha = (\theta, h) \in \widehat{\mathcal{A}}_n$ and let $p_u(\alpha) = E[L(h)]$ with $\alpha = (\theta, h) \in \mathcal{A}_+$. Now note that for any $\varepsilon > 0$,

$$P(d(\widehat{\mathcal{P}}_n, \mathcal{P}_+) > \varepsilon) \leq P(\rho(\widehat{\mathcal{P}}_n|\mathcal{P}_+) > \varepsilon) + P(\rho(\mathcal{P}_+|\widehat{\mathcal{P}}_n) > \varepsilon) \tag{38}$$

from the definition of $d(\cdot,\cdot)$. Let $\epsilon_\delta = P(d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) > \delta)$ and consider

$$\begin{aligned} &P(\rho(\widehat{\mathcal{P}}_n|\mathcal{P}_+) > \varepsilon) \leq P(\rho(\widehat{\mathcal{P}}_n|\mathcal{P}_+) > \varepsilon, d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \leq \delta) + \epsilon_\delta \\ &= P(\sup_{\alpha \in \widehat{\mathcal{A}}_n} \inf\{|p_{u,n}(\alpha) - p_u(\alpha_+)| : \alpha_+ \in \mathcal{A}_+\} > \varepsilon, d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \leq \delta) + \epsilon_\delta \end{aligned} \tag{39}$$

where the equality holds by definitions of $\rho(\cdot|\cdot)$, $\widehat{\mathcal{P}}_n$, and $\mathcal{P}_+$. Now consider that if $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \leq \delta$, then $\rho(\widehat{\mathcal{A}}_n|\mathcal{A}_+) \leq \delta$ and for any $\alpha \in \widehat{\mathcal{A}}_n$, there exists $\alpha_{++} \in \mathcal{A}_+$ such that $\|\alpha - \alpha_{++}\|_s \leq \delta$. We let $\alpha = (\theta, h)$ and $\alpha_{++} = (\theta_{++}, h_{++})$. It follows that for any $\alpha \in \widehat{\mathcal{A}}_n$ such that $\|\alpha - \alpha_{++}\|_s \leq \delta$, we have

$$\begin{aligned} &p_{u,n}(\alpha) - p_u(\alpha_{++}) \\ &= \tfrac{1}{n}\textstyle\sum_{i=1}^n (L(h(Z_i)) - L(h_{++}(Z_i))) + \tfrac{1}{n}\sum_{i=1}^n (L(h_{++}(Z_i)) - E[L(h_{++}(Z_i))]) \\ &= \tfrac{1}{n}\textstyle\sum_{i=1}^n L(\widetilde{h}(Z_i))(1 - L(\widetilde{h}(Z_i)))(h(Z_i) - h_{++}(Z_i)) + \tfrac{1}{n}\sum_{i=1}^n (L(h_{++}(Z_i)) - E[L(h_{++}(Z_i))]) \\ &\leq \tfrac{1}{4}\|h - h_{++}\|_{\mathcal{H}} + o_p(1) \leq \tfrac{1}{4}\|\alpha - \alpha_{++}\|_s + o_p(1) \leq \delta \text{ for sufficiently large } n \geq \exists N, \end{aligned}$$

where the second equality is obtained applying the mean value theorem and the first inequality holds since $L(1-L) \leq 1/4$ uniformly and since we bound the second right-hand side (RHS) term of the second equality by $o_p(1)$ applying the LLN ($\{Z_i\}_{i=1}^n$ are iid and $|L(h)| < 1$ uniformly). From this result, we have $\sup_{\alpha \in \widehat{\mathcal{A}}_n} \inf\{|p_{u,n}(\alpha) - p_u(\alpha_+)| : \alpha_+ \in \mathcal{A}_+\} \leq \sup_{\alpha \in \widehat{\mathcal{A}}_n, \|\alpha - \alpha_{++}\|_s \leq \delta} |p_{u,n}(\alpha) - p_u(\alpha_{++})| \leq \delta$ for all sufficiently large $n \geq \exists N$. From this, it follows that for sufficient large n,

$$P(\sup_{\alpha \in \widehat{\mathcal{A}}_n} \inf\{|p_{u,n}(\alpha) - p_u(\alpha_+)| : \alpha_+ \in \mathcal{A}_+\} > \varepsilon, d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \leq \delta) \leq \varepsilon \tag{40}$$

and thus from (39) and (40), we obtain $P(\rho(\widehat{\mathcal{P}}_n|\mathcal{P}_+) > \varepsilon) \leq \varepsilon + P(d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) > \delta)$. An analogous argument provides $P(\rho(\mathcal{P}_+|\widehat{\mathcal{P}}_n) > \varepsilon) \leq \varepsilon + P(d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) > \delta)$. Combining these two results with (38), we conclude $P(d(\widehat{\mathcal{P}}_n, \mathcal{P}_+) > \varepsilon) \leq 2\varepsilon + 2P(d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) > \delta)$. This proves Proposition 4.1 since $\varepsilon > 0$ is arbitrary and $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) = o_p(1)$.

B Large Sample Theory for CI

B.1 High-level Assumptions and Primitive Conditions

Here we provide a semiparametric version of high-level assumptions given in ABJ to justify the CI's asymptotically. We also provide sets of sufficient conditions that satisfy some of such high-level assumptions. Define $\mathcal{H}_\delta \equiv \{h \in \mathcal{H} : \|h(\cdot,\theta) - h_0(\cdot,\theta)\|_{\mathcal{H}} \leq \delta,\ \theta \in \Theta\}$ for some small $\delta > 0$ and let $\mathcal{A}_\delta \equiv \Theta \times \mathcal{H}_\delta$. This section provides the high-level assumptions that justify the CI's introduced in Section 5. Let

$$\widehat{\nu}_n(j,\gamma,\theta,h) = \sqrt{n}(\widehat{c}_n(j,\gamma,\theta,h) - c_0(j,\gamma,\theta,h)) \text{ and} \tag{41}$$

$$\widehat{Z}_n(j,m,\theta,h) = \sqrt{n}(c_0(j,\widehat{\gamma}_{n,j,m},\theta,h) - c_0(j,\gamma_{0,j,m},\theta,h)). \tag{42}$$

Viewed as a function of (j,γ,θ,h), $\widehat{\nu}_n(j,\gamma,\theta,h)$ is a stochastic process on $\mathcal{I}_J \times \Gamma_{all} \times \mathcal{A}_\delta$. Under suitable conditions, $\widehat{\nu}_n(j,\gamma,\theta,h)$ converges weakly to a mean zero Gaussian process $\nu_0(j,\gamma,\theta,h)$ on $\mathcal{I}_J \times \Gamma_{all} \times \mathcal{A}_\delta$. The covariance function of $\nu_0(\cdot,\cdot,\cdot,\cdot)$ is given by

$$\begin{aligned} &V_0((j_1,\gamma_1,\theta_1,h_1),(j_2,\gamma_2,\theta_2,h_2)) \equiv Cov(\nu_0(j_1,\gamma_1,\theta_1,h_1),\nu_0(j_2,\gamma_2,\theta_2,h_2)) \\ &= E\left[\begin{array}{c} \left(\left(P(y^{(j_1)}|X,\theta_1,h_1) - \mathbf{1}[Y=y^{(j_1)}]\right)q_{\gamma_1}(X) - c_0(j_1,\gamma_1,\theta_1,h_1)\right) \\ \times\left(\left(P(y^{(j_2)}|X,\theta_2,h_2) - \mathbf{1}[Y=y^{(j_2)}]\right)q_{\gamma_2}(X) - c_0(j_2,\gamma_2,\theta_2,h_2)\right) \end{array}\right]. \end{aligned} \tag{43}$$

for $(\theta_1,h_1),(\theta_2,h_2) \in \mathcal{A}_\delta$. Now let $\widehat{Z}_n(\theta,h)$ denote the $\sum_{j=1}^J \mathcal{M}_j \times 1$ column vector whose elements are $\{\widehat{Z}_n(j,m,\theta,h) : (j,m) \in \mathcal{I}_{J,\mathcal{M}}\}$ such that $\widehat{Z}_n(1,1,\theta,h)$ is the first element and $\widehat{Z}_n(1,2,\theta,h)$ is the second element, etc. At last, let $\Rightarrow$ denote weak convergence of a sequence of stochastic processes. The following assumptions extend the assumptions in ABJ allowing for infinite dimensional parameters.

Assumption B.1 *$\widehat{\nu}_n(\cdot,\cdot,\cdot,\cdot) \Rightarrow \nu_0(\cdot,\cdot,\cdot,\cdot)$, where $\nu_0(\cdot,\cdot,\cdot,\cdot)$ is a mean zero Gaussian process indexed by $(j,\gamma,\theta,h) \in \mathcal{I}_J \times \Gamma_{all} \times \mathcal{A}_\delta$ with bounded and continuous sample paths a.s. (with respect to $\|\cdot\|$ on Γ_{all} and the metric $\|\cdot\|_s$ on $\mathcal{A}_\delta$) with covariance function $V_0(\cdot)$ defined in (43).*

We note that the following stochastic equicontinuity condition is sufficient for Assumption B.1:

$$\sup_{\|(\theta',h')-(\theta,h)\|_s \leq \delta_n, \|\gamma'-\gamma\| \leq \delta_n} \left|\widehat{\nu}_n(j,\gamma',\theta',h') - \widehat{\nu}_n(j,\gamma,\theta,h)\right| = o_p(1), \text{ for any given } (\gamma,\theta,h) \in \Gamma_{all} \times \mathcal{A}_\delta \tag{44}$$

for any positive sequence δ_n tending to zero.

Recall that $\mathcal{A}_\delta$ is compact (with respect to $\|\cdot\|_s$). When X is discrete, then it is obvious Γ_{all} is finite. When X is continuous, we construct γ's such that they have non-empty interior. Then, Γ_{all} is still a finite set since the number of all subsets (with nonempty interior) of $\mathbb{S}(X)$ is finite due to the compactness of $\mathbb{S}(X)$ (any compact set is totally bounded). Thus, Γ_{all} is totally bounded with respect to $\|\cdot\|$. Therefore, $\left((\mathcal{A}_\delta,\Gamma_{all}),(\|\cdot\|_s,\|\cdot\|)\right)$ is a totally bounded pseudometric space. It is not difficult to show the finite dimensional (fidi) convergence holds, i.e, all the finite subsets $((\theta_1,h_1,\gamma_1),\ldots,(\theta_J,h_J,\gamma_J))$ of $(\mathcal{A}_\delta,\Gamma_{all})$, $(\widehat{\nu}_n(j,\gamma_1,\theta_1,h_1),\ldots,\widehat{\nu}_n(j,\gamma_J,\theta_J,h_J))'$ converge in distribution. Therefore, as long as the condition (44) holds, Assumption B.1 holds by the weak convergence theorem (see Pollard (1990) or van der Vaart and Wellner (1996)). (44) can be proved under suitable conditions similarly with Chen, Linton, and van Keilegom (2003) as in the following lemma.

Lemma B.1 *Suppose Assumptions 4.1 and 4.4 hold. Further suppose that (a) Θ is a compact subset of $\mathbb{R}^{d_\theta}$ and $\int_0^\infty \sqrt{\log N(\varepsilon,\mathcal{H},\|\cdot\|_s)}d\varepsilon < \infty$; and that (b) (Lipschitz Condition) (i) For $P(Y=y^{(j)}|X,\theta,h)$, $j=1,\ldots,J$, the pathwise derivative at the direction $[(\widetilde{\theta},\widetilde{h})-(\theta,h)]$ exists for all $(\widetilde{\theta},\widetilde{h}),(\theta,h) \in \Theta \times \mathcal{H}_\delta$ and hence for $\exists M_1(j,X,\theta,h) \equiv \frac{dP(Y=y^{(j)}|X,\theta,h)}{d\theta'}$ and $\exists m_2(\cdot,\cdot,\cdot,\cdot)$, we have $\frac{dP(Y=y^{(j)}|X,\overline{\theta},\overline{h})}{d(\theta,h)}[(\widetilde{\theta},\widetilde{h})-(\theta,h)] = M_1(j,X,\overline{\theta},\overline{h})\left(\widetilde{\theta}-\theta\right) + \sum_{v=1}^{\boldsymbol{v}} m_{2v}(j,X,\overline{\theta},\overline{h})\left(\widetilde{h}_v - h_v\right)$; (ii) $\sup_{(\theta,h)\in\Theta\times\mathcal{H}_\delta}|M_1(j,X,\theta,h)| \leq C_1(X) < \infty$ and $\sup_{(\theta,h)\in\Theta\times\mathcal{H}_\delta}|m_{2v}(j,X,\theta,h)| \leq C_{2v}(X) < \infty$ for all $j=1,\ldots,J$ and $v=1,\ldots,\boldsymbol{v}$. Then, Condition (44) holds.*

Assumption B.2 *$\widehat{Z}_n(\theta,h) \underset{d}{\rightarrow} Z_0(\theta,h)$, where $Z_0(\cdot,\cdot)$ is a mean zero Gaussian process indexed by $(\theta,h) \in \mathcal{A}_\delta$ with bounded and continuous sample paths a.s. (with respect to the metric $\|\cdot\|_s$) and the convergence holds jointly with that in Assumption B.1 with the joint limit being Gaussian.*

Assumption B.3 *(i)* $\sup_{(j,\gamma,\theta,h)\in\mathcal{I}_J\times\Gamma_{all}\times\mathcal{A}_\delta}|\widehat{w}_n(j,\gamma,\theta,h)-w_0(j,\gamma,\theta,h)|\underset{p}{\rightarrow}0$ *for some non-random positive functional* w_0 *on* $\mathcal{I}_J\times\Gamma_{all}\times\mathcal{A}_\delta$ *that is bounded and bounded away from zero. (ii)* $w_0(j,\gamma,\theta,h)$ *is continuous in* (γ,θ,h) *(with respect to the product of* $\|\cdot\|$ *on* Γ_{all} *and the metric* $\|\cdot\|_s$ *on* $\mathcal{A}_\delta$*) at* $(\gamma_{0,j,m},\theta,h)$, $\forall(\theta,h)\in\mathcal{A}_\delta$, $\forall(j,m)\in\mathcal{I}_{J,\mathcal{M}}$.

To develop the asymptotics, we need to define the population analogues of $\widehat{\alpha}_{n,U}$ and $\widehat{\alpha}_{n,L}$ defined in (16) replacing $\widehat{\mathcal{A}}_{n,U}$ and $\widehat{\beta}_{n,U}$ with $\mathcal{A}_+$ and $\beta_{n,U}$, respectively. We let[22]

$$\alpha_{n,U}=\text{argmin}\left\{\|\alpha\|_s:\alpha\in\mathcal{A}_+,\beta_n(\alpha)=\beta_{n,U}\right\}\tag{45}$$

and define $\alpha_{n,L}$ analogously with $\beta_{n,L}$ in place of $\beta_{n,U}$.

Assumption B.4 *(i)* $\beta_n(\cdot)\underset{p}{\rightarrow}\beta_0(\cdot)$ *uniformly over* $\alpha\in\mathcal{A}_\delta$ *for some non-random continuous functional* $\beta_0(\cdot)$ *on* $\mathcal{A}_\delta$*. (ii)* $\|\alpha_{n,U}-\alpha_{+,U}\|_s\underset{p}{\rightarrow}0$ *and* $\|\alpha_{n,L}-\alpha_{+,L}\|_s\underset{p}{\rightarrow}0$*, where* $\alpha_{+,U}=\text{argmin}\left\{\|\alpha\|_s:\alpha\in\mathcal{A}_+,\beta_0(\alpha)=\beta_{+,U}\right\}$, $\beta_{+,U}=\sup\left\{\beta_0(\alpha):\alpha\in\mathcal{A}_+\right\}$, $\beta_{+,L}$ *is defined analogously with* sup *replaced by* inf*, and* $\alpha_{+,L}$ *is defined with* $\beta_{+,U}$ *replaced by* $\beta_{+,L}$.

Note that if $\beta_n(\cdot)$ is non-random and does not depend on n, Assumption B.4 immediately holds by construction. Now define

$$\mathcal{B}_{n,U}=\left\{(j,m)\in\mathcal{I}_{J,\mathcal{M}}:c_0(j,\gamma_{0,j,m},\alpha_{n,U})=0\right\},\ \mathcal{B}_{+,U}=\left\{(j,m)\in\mathcal{I}_{J,\mathcal{M}}:c_0(j,\gamma_{0,j,m},\alpha_{+,U})=0\right\},\tag{46}$$

and define $\mathcal{B}_{n,L}$ & $\mathcal{B}_{+,L}$ analogously replacing $\alpha_{n,U}$ and $\alpha_{+,U}$ with $\alpha_{n,L}$ and $\alpha_{+,L}$, respectively. We assume

Assumption B.5 *(i)* $\widehat{\beta}_{n,U}-\beta_{n,U}\underset{p}{\rightarrow}0$ *and* $\widehat{\beta}_{n,L}-\beta_{n,L}\underset{p}{\rightarrow}0$*; (ii)* $P(\widehat{\mathcal{B}}_{n,U}\subseteq\mathcal{B}_{n,U}\subseteq\mathcal{B}_{+,U})\rightarrow1$ *and* $P(\widehat{\mathcal{B}}_{n,L}\subseteq\mathcal{B}_{n,L}\subseteq\mathcal{B}_{+,L})\rightarrow1$.

Theorem 4.1 provides sufficient conditions for Assumption B.5 (i). ABJ note that by allowing the estimated binding constraints sets $\widehat{\mathcal{B}}_{n,U}$ and $\widehat{\mathcal{B}}_{n,L}$ to be smaller than the population versions $\mathcal{B}_{n,U}$ and $\mathcal{B}_{n,L}$[23], a researcher can consider more constraints in the estimation stage then in the CI construction stage. Lastly, we assume that when employing the CI for β_0, the critical values $\lambda^*_{n,U}(j,m,a)$ and $\lambda^*_{n,L}(j,m,a)$ defined in (23) converge in probability to non-negative constants $\lambda_{0,U}(j,m,a)$ and $\lambda_{0,L}(j,m,a)$ that satisfy

$$P\begin{pmatrix}\nu_0(j,\gamma_{0,j,m},\alpha_{+,U})+Z_0(j,m,\alpha_{+,U})+w_0(j,\gamma_{0,j,m},\alpha_{+,U})\lambda_{0,U}(j,m,a)\geq0\text{ for }(j,m)\in\mathcal{B}_{+,U}\\ \&\ \nu_0(j,\gamma_{0,j,m},\alpha_{+,U})+Z_0(j,m,\alpha_{+,U})+w_0(j,\gamma_{0,j,m},\alpha_{+,U})\lambda_{0,L}(j,m,a)\geq0\text{ for }(j,m)\in\mathcal{B}_{+,L}\end{pmatrix}=1-a\tag{47}$$

and the same condition holds with U and L interchanged.

Assumption B.6 *For the CI of the true value* β_0, $\lambda^*_{n,U}(j,m,a)\underset{p}{\rightarrow}\lambda_{0,U}(j,m,a)\geq0$ *for all* $(j,m)\in\mathcal{B}_{+,U}$ *and* $\lambda^*_{n,L}(j,m,a)\underset{p}{\rightarrow}\lambda_{0,L}(j,m,a)\geq0$ *for all* $(j,m)\in\mathcal{B}_{+,L}$ *where* $\lambda_{0,U}(j,m,a)$ *and* $\lambda_{0,L}(j,m,a)$ *satisfy (47).*

Suppose that

$$\sqrt{n}\left(\frac{\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})-c_0(j,\gamma_{0,j,m},\widehat{\alpha}_{n,U})}{\widehat{w}_n\left(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U}\right)}\right)\underset{d}{\rightarrow}\frac{\nu_0(j,\gamma_{0,j,m},\alpha_{+,U})+Z_0(j,m,\alpha_{+,U})}{w_0\left(j,\gamma_{0,j,m},\alpha_{+,U}\right)}\tag{48}$$

and suppose the same condition holds with U replaced by L. Then, Assumption B.6 will be satisfied if

$$\sqrt{n}\left(\frac{\widehat{c}^*_n(j,\widehat{\gamma}^*_{n,j,m},\widehat{\alpha}_{n,U})-\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})}{w^*_n\left(j,\widehat{\gamma}^*_{n,j,m},\widehat{\alpha}_{n,U}\right)}\right)\underset{d}{\rightarrow}\frac{\nu_0(j,\gamma_{0,j,m},\alpha_{+,U})+Z_0(j,m,\alpha_{+,U})}{w_0\left(j,\gamma_{0,j,m},\alpha_{+,U}\right)}\tag{49}$$

in P^*-probability and the same condition holds with U replaced by L. Lemma B.2 provides sufficient conditions for (48) and Lemma B.3 provides sufficient conditions for (49). Analogous sufficient conditions for (48) and (49) with U replaced by L can be found in Lemma B.2 and Lemma B.3 with L in replace of U.

[22] Here we assume that $\alpha_{n,U}$ is unique. If it is not unique, then we select one of those functions that satisfy (45) according to a certain criterion.

[23] Note that the use of fewer constraints cannot reduce the coverage probability of the CI.

Lemma B.2 *Suppose that $\mathcal{A}_+$ satisfies the condition (5), that $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \underset{p}{\rightarrow} 0$, and that (i) $\|\widehat{\alpha}_{n,U} - \alpha_{n,U}\|_s = o_p(1)$; (ii) (44) holds; (ii) Assumptions B.2, B.3, B.4, and 4.7 hold. Then, (48) is satisfied.*

Note that Assumptions B.2, B.3 and 4.7 can be verified for a particular choice of $\widehat{\Gamma}_n$ and a weight function $\widehat{w}_n(\cdot)$. Thus, they are directly assumed in this paper. To prove Lemma B.3, we need the following condition:

$$\sup_{\|\alpha'-\alpha\|_s \leq \delta_n, \|\gamma'-\gamma\| \leq \delta_n} \left|\widehat{c}_n^*(j,\gamma',\alpha') - \widehat{c}_n(j,\gamma',\alpha') - (\widehat{c}_n^*(j,\gamma,\alpha) - \widehat{c}_n(j,\gamma,\alpha))\right| = o_{P^*}(n^{-1/2}) \tag{50}$$

for any given $(\gamma, \alpha) \in \Gamma_{all} \times \mathcal{A}_\delta$. This is a bootstrap version of the stochastically equicontinuity condition of (44) and will be satisfied under the same sufficient conditions for (44).

Lemma B.3 *Suppose that $\mathcal{A}_+$ satisfies the condition (5), that $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \rightarrow 0$ a.s., and that (i) $\|\widehat{\alpha}_{n,U} - \alpha_{n,U}\|_s = o(1)$ a.s.; (ii) $\sup_{(j,\gamma,\alpha)\in\mathcal{I}_J\times\Gamma_{all}\times\mathcal{A}_\delta} |\widehat{w}_n^*(j,\gamma,\alpha) - \widehat{w}_n(j,\gamma,\alpha)| = o_{P^*}(1)$; (iii) (50) holds; (iv) (44) and Assumptions B.3, B.4, and 4.7 hold with "in probability" replaced by "almost surely". Then, (49) holds in P^*-probability.*

B.2 Asymptotics for Confidence Interval

The proof of Theorem 5.1 and the proofs of lemmas in Appendix B use the following lemma, which extends Lemma 5 of ABJ to the semiparametric case.

Lemma B.4 *Suppose Assumptions B.1-B.3, B.6, 4.2, 4.3 (i), 4.4 (i), and 4.7 for all $(j,m) \in \mathcal{I}_{J,\mathcal{M}}$ hold. Then, for any $\alpha \in \mathcal{A}_\delta$ such that $\|\alpha - \alpha_{+,U}\|_s \rightarrow 0$, we have (i) $\widehat{\nu}_n(j, \widehat{\gamma}_{n,j,m}, \alpha) \underset{d}{\rightarrow} \nu_0(j, \gamma_{0,j,m}, \alpha_{+,U})$; (ii) $\widehat{Z}_n(j,m,\alpha) \underset{d}{\rightarrow} Z_0(j,m,\alpha_{+,U})$; (iii) $\widehat{w}_n(j, \widehat{\gamma}_{n,j,m}, \alpha) \underset{d}{\rightarrow} w_0(j, \gamma_{0,j,m}, \alpha_{+,U})$; (iv) $\widehat{\nu}_n(j, \widehat{\gamma}_{n,j,m}, \alpha) + \widehat{Z}_n(j,m,\alpha) + \widehat{w}_n(j, \widehat{\gamma}_{n,j,m}, \alpha)\lambda_{n,U}^*(j,m,a) \underset{d}{\rightarrow} \nu_0(j, \gamma_{0,j,m}, \alpha_{+,U}) + Z_0(j,m,\alpha_{+,U}) + w_0(j, \gamma_{0,j,m}, \alpha_{+,U})\lambda_{0,U}(j,m,a)$; and (v) the results of parts (i)-(iv) hold with U replaced by L and all the convergence results of the lemma hold jointly.*

Proof. Combining Assumptions B.1 and 4.7, we obtain for any $\alpha \in \mathcal{A}_\delta$ s.t. $\|\alpha - \alpha_{+,U}\|_s \rightarrow 0$, $(\widehat{\nu}_n(\cdot,\cdot,\cdot), \widehat{\Gamma}_n, \alpha) \Rightarrow (\nu_0(\cdot,\cdot,\cdot), \Gamma_0, \alpha_{+,U})$ as processes indexed by $(j,\gamma,\alpha) \in \mathcal{I}_J \times \Gamma_{all} \times \mathcal{A}_\delta$ and this convergence is joint with that in Assumption B.2. Note that the function $g(\nu(\cdot,\cdot,\cdot), \Gamma, \alpha) = \nu(j, \gamma_{j,m}, \alpha)$ is continuous at $(\nu_0(\cdot,\cdot,\cdot), \Gamma_0, \alpha_{+,U})$ because $\nu_0(j,\cdot,\cdot)$ has continuous sample paths a.s. with respect to the product of the $\|\cdot\|$ norm and the $\|\cdot\|_s$ metric. Thus, applying the continuous mapping theorem (e.g., see Pollard (1984)), for any $\alpha \in \mathcal{A}_\delta$ such that $\|\alpha - \alpha_{+,U}\|_s \rightarrow 0$, we find $\widehat{\nu}_n(j, \widehat{\gamma}_{n,j,m}, \alpha) \underset{d}{\rightarrow} \nu_0(j, \gamma_{0,j,m}, \alpha_{+,U})$, which proves part (i).

Similarly with part (i), we see part (ii) holds by Assumptions B.2 from the continuous mapping theorem.

Next, we prove part (iii). Using the triangle inequality and Assumption B.3, we have for any $\alpha \in \mathcal{A}_\delta$ such that $\|\alpha - \alpha_{+,U}\|_s \rightarrow 0$,

$$\begin{aligned}
&\left|\widehat{w}_n\left(j, \widehat{\gamma}_{n,j,m}, \alpha\right) - w_0\left(j, \gamma_{0,j,m}, \alpha_{+,U}\right)\right| \\
\leq &\left|\widehat{w}_n\left(j, \widehat{\gamma}_{n,j,m}, \alpha\right) - w_0\left(j, \widehat{\gamma}_{n,j,m}, \alpha\right)\right| + \left|w_0\left(j, \widehat{\gamma}_{n,j,m}, \alpha\right) - w_0\left(j, \gamma_{0,j,m}, \alpha_{+,U}\right)\right| \\
\leq &\sup_{(j,\gamma,\alpha)\in\mathcal{I}_J\times\Gamma_{all}\times\mathcal{A}_\delta} \left|\widehat{w}_n\left(j,\gamma,\alpha\right) - w_0\left(j,\gamma,\alpha\right)\right| + \left|w_0\left(j, \widehat{\gamma}_{n,j,m}, \alpha\right) - w_0\left(j, \gamma_{0,j,m}, \alpha_{+,U}\right)\right| \underset{p}{\rightarrow} 0
\end{aligned}$$

where the first RHS term in the second inequality goes to zero by Assumption B.3 (i) and the second term goes to zero by the continuity assumed in Assumption B.3 (ii). Combining parts (i)-(iii) of the lemma and Assumption B.6 proves part (iv). ■

For the proof of Theorem 5.1, we also need

Lemma B.5 *Suppose $\widetilde{\beta}_{l,n,\cdot}^{(3)}$ has a well-defined conditional cdf (conditional on $\beta_{n,\cdot}$) whose derivative is uniformly bounded over its support. Then, for any small $\eta,\epsilon > 0$, we have $P(\beta_{n,U} \leq \widetilde{\beta}_{l,n,U}^{(3)} + \eta) - P(\beta_{n,U} \leq \widetilde{\beta}_{l,n,U}^{(3)}) < \epsilon$ and $P(\beta_{n,L} \geq \widetilde{\beta}_{l,n,L}^{(3)} - \eta) - P(\beta_{n,L} \geq \widetilde{\beta}_{l,n,L}^{(3)}) < \epsilon$ uniformly over $\beta_{n,U}$ and $\beta_{n,L}$.*

Proof. We treat $\beta_{n,L}$ non-random without loss of generality. By the mean value theorem, for some $\varsigma \in [0,1]$, we have $P(\beta_{n,U} \leq \widetilde{\beta}_{l,n,U}^{(3)} + \eta) - P(\beta_{n,U} \leq \widetilde{\beta}_{l,n,U}^{(3)}) = 1 - G(\beta_{n,U} - \eta) - (1 - G(\beta_{n,U})) = \dot{G}\left(\beta_{n,U} - \varsigma \cdot \eta\right)\eta$ where $G(\cdot)$ is the cdf of $\widetilde{\beta}_{l,n,U}^{(3)}$ and $\dot{G}(\cdot)$ is its derivative. Thus, the first claim follows as long as $\dot{G}(\cdot)$ is uniformly bounded. The second claim can be proved similarly. ■

B.2.1 Proof of Theorem 5.1

By Assumption B.4 (i), $\beta_0 = \beta_n(\alpha_0) \underset{p}{\to} \beta_0(\alpha_0)$. We let $\beta_{0,0} = \beta_0(\alpha_0)$ denote the asymptotic true value. The following cases are considered separately: (i) $\beta_{+,L} < \beta_{0,0} < \beta_{+,U}$, (ii) $\beta_{+,L} < \beta_{0,0} = \beta_{+,U}$, (iii) $\beta_{+,L} = \beta_{0,0} < \beta_{+,U}$, and (iv) $\beta_{+,L} = \beta_{0,0} = \beta_{+,U}$. The proofs are given for all three alternative CI's. We let $\widetilde{\beta}_{n,L}$ and $\widetilde{\beta}_{n,U}$ denote the generic lower bound and the generic upper bound for three alternative CI's, which suppress l in Alternative CI3 for notational convenience.

Case (i): We have $\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha) \geq \widehat{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha)$ for all (j,m,α), since $\widehat{w}_n(j,\gamma,\alpha) > 0$ and $\lambda^*_{n,U}(j,m,a) \geq 0$. For all the three alternative CI's, this implies that $\widetilde{\beta}_{n,U} \geq \widehat{\beta}_{n,U}$ by constructions of $\widehat{\beta}_{n,U}$ and $\widetilde{\beta}_{n,U}$ in (14) and (19), (20), or (21) (note $\widehat{\mathcal{A}}_n \subseteq \Theta \times \widehat{\mathcal{H}}_{l,\delta_n} \subseteq \Theta \times \widehat{\mathcal{H}}_{\delta_n} \subseteq \Theta \times \mathcal{H}$ by construction), respectively. Combining this with Theorem 4.1 and Assumption B.4 (i) gives $\widetilde{\beta}_{n,U} - \beta_0 \geq \widehat{\beta}_{n,U} - \beta_0 \underset{p}{\to} \beta_{+,U} - \beta_{0,0} > 0$ and $P(\beta_0 \leq \widetilde{\beta}_{n,U}) \to 1$. By an analogous argument, we can show that $P(\beta_0 \geq \widetilde{\beta}_{n,L}) \to 1$, which establishes the result of Theorem 5.1 for case (i).

Case (ii): From $\beta_{0,0} > \beta_{+,L}$ and the same argument as above, it follows that $P(\beta_0 \geq \widetilde{\beta}_{n,L}) \to 1$. It remains to show that $\liminf_{n\to\infty} P(\beta_0 \leq \widetilde{\beta}^{(1)}_{n,U}) \geq 1-a$ for Alternative CI1, $\liminf_{n\to\infty} P(\beta_0 \leq \widetilde{\beta}^{(2)}_{n,U}) \geq 1-a$ for Alternative CI2, and $\liminf_{l,n\to\infty, l\geq n} P(\beta_0 \leq \widetilde{\beta}^{(3)}_{l,n,U}) \geq 1-a$ for Alternative CI3. We start with Alternative CI1. From definition of $\alpha_{n,U}$ in (45), we have $\beta_n(\alpha_{n,U}) = \beta_{n,U}$. Also note that if $\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}$, then $\widetilde{\beta}^{(1)}_{n,U}$ cannot be smaller than $\beta_{n,U}$ by constructions of $\widetilde{\beta}^{(1)}_{n,U}$ and $\beta_{n,U}$ ($\alpha_{n,U}$ becomes an element of the set to which we take the sup operator to obtain $\widetilde{\beta}^{(1)}_{n,U}$). Moreover, $\beta_0 \leq \beta_{n,U}$ by definition of $\beta_{n,U}$ in (13) and the fact that $\alpha_0 \in \mathcal{A}_+$. Combining these results, we obtain

$$P(\beta_0 \leq \widetilde{\beta}^{(1)}_{n,U}) \geq P(\beta_{n,U} \leq \widetilde{\beta}^{(1)}_{n,U}) \geq P(\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}) \tag{51}$$

for all $n \geq \exists N$. Now note that

$$\begin{aligned}
&\liminf_{n\to\infty} P(\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}) \\
&= \liminf_{n\to\infty} P(\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq c_0(j,\gamma_{0,j,m},\alpha_{n,U}) \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U} \text{ and } \widehat{\mathcal{B}}_{n,U} \subseteq \mathcal{B}_{n,U}) \\
&\quad + \liminf_{n\to\infty} P(\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U} \text{ and } \widehat{\mathcal{B}}_{n,U} \nsubseteq \mathcal{B}_{n,U}) \\
&\geq \liminf_{n\to\infty} P(\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq c_0(j,\gamma_{0,j,m},\alpha_{n,U}) \ \forall (j,m) \in \mathcal{B}_{n,U})
\end{aligned} \tag{52}$$

where the equality holds because $c_0(j,\gamma_{0,j,m},\alpha_{n,U}) = 0 \ \forall (j,m) \in \mathcal{B}_{n,U}$ by definition of $\mathcal{B}_{n,U}$ in (46), the inequality holds by Assumption B.5 (ii) and the fact that a set can not be larger when it is defined using more restrictions.

Let $Q_{n,U} = P\left(\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq c_0(j,\gamma_{0,j,m},\alpha_{n,U}) \ \forall (j,m) \in \mathcal{B}_{n,U}\right)$. This can be rewritten as $Q_{n,U} = P\left(\widehat{\nu}_n(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) + \widehat{Z}_n(j,m,\alpha_{n,U}) + \widehat{w}_n(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U})\lambda^*_{n,U}(j,m,a) \geq 0 \ \forall (j,m) \in \mathcal{B}_{n,U}\right)$ using the definitions of $\widetilde{c}_{n,U}(j,\gamma,\alpha)$, $\widehat{\nu}_n(j,\gamma,\alpha)$, and $\widehat{Z}_n(j,m,\alpha)$. From Lemma B.4 (iv) (note $\alpha_{n,U} \in \mathcal{A}_+ \subset \mathcal{A}_\delta$) and Assumption B.5 (ii), it follows that

$$\liminf_{n\to\infty} Q_{n,U} \geq P\left(\nu_0(j,\gamma_{0,j,m},\alpha_{+,U}) + Z_0(j,m,\alpha_{+,U}) + w_0(j,\gamma_{0,j,m},\alpha_{+,U})\lambda_{0,U}(j,m,a) \geq 0 \ \forall (j,m) \in \mathcal{B}_{+,U}\right). \tag{53}$$

Note that the strict inequality in (53) is allowed since Assumption B.5 (ii) allows $\mathcal{B}_{n,U}$ to be a strict subset of $\mathcal{B}_{+,U}$ wpa1. If $\beta_n(\cdot)$ is non-random and does not depend on n, we have $\mathcal{B}_{n,U} = \mathcal{B}_{+,U}$ by definition and hence, (53) holds with equality. Now note that by definition of $\lambda_{0,U}(j,m,a)$ in (47), the RHS of (53) is greater than or equal to $1-a$. This completes the proof of Theorem 5.1 for case (ii) with Alternative CI1.

Now we turn to Alternative CI2. Note that $\Theta_+ \times \mathcal{H}_+$ is included in $\Theta \times \widehat{\mathcal{H}}_\delta$ wpa1 because $d(\widehat{\mathcal{A}}_n, \mathcal{A}_+) \to 0$ and $\widehat{\mathcal{A}}_n \subseteq \Theta \times \widehat{\mathcal{H}}_\delta$. Thus, if $\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}$, then $\widetilde{\beta}^{(2)}_{n,U}$ cannot be smaller than $\beta_{n,U}$ by constructions of $\widetilde{\beta}^{(2)}_{n,U}$ and $\beta_{n,U}$ wpa1 since wpa1, $\alpha_{n,U} \in \Theta \times \widehat{\mathcal{H}}_{\delta_n}$. Similarly with Alternative CI1, we obtain

$$\begin{aligned}
&P(\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}) \\
&= P\left(\left\{\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}\right\} \text{ and } \left\{\alpha_{n,U} \in \Theta \times \widehat{\mathcal{H}}_{\delta_n}\right\}\right) \\
&\quad + P\left(\left\{\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}\right\} \text{ and } \left\{\alpha_{n,U} \notin \Theta \times \widehat{\mathcal{H}}_{\delta_n}\right\}\right) \\
&\leq P(\beta_{n,U} \leq \widetilde{\beta}^{(2)}_{n,U}) + P(\alpha_{n,U} \notin \Theta \times \widehat{\mathcal{H}}_{\delta_n}) \leq P(\beta_0 \leq \widetilde{\beta}^{(2)}_{n,U}) + o_p(1)
\end{aligned} \tag{54}$$

where the last inequality holds by the same reason with Alternative CI1 and $\alpha_{n,U} \in \Theta \times \widehat{\mathcal{H}}_{\delta_n}$ wpa1. The remaining proof exactly follows that of Alternative CI1 and thus this completes the proof of Theorem 5.1 for case (ii) with Alternative CI2.

Now we turn to Alternative CI3. Recall that $\widetilde{\beta}^{(3)}_{l,n,U} = \sup\{\beta_n(\alpha) : \alpha \in \Theta \times \widehat{\mathcal{H}}_{l,\delta_n}, \widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha) \geq 0, \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}\}$ and $\widetilde{\beta}^{(3)}_{\infty,n,U} = \sup\left\{\beta_n(\alpha) : \alpha \in \Theta \times \widehat{\mathcal{H}}_{\delta_n}, \widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha) \geq 0, \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}\right\}$. Then, by construction of $\widetilde{\beta}_{l,n,U}$, we have $\widetilde{\beta}^{(3)}_{l,n,U} \leq \widetilde{\beta}^{(3)}_{l+1,n,U} \leq \widetilde{\beta}^{(3)}_{\infty,n,U}$ for all $l \geq 1$ since $\Theta \times \widehat{\mathcal{H}}_{l,\delta_n} \subseteq \Theta \times \widehat{\mathcal{H}}_{l+1,\delta_n} \subseteq \Theta \times \widehat{\mathcal{H}}_{\delta_n}$[24] for all $l \geq 1$ and thus $\widetilde{\beta}^{(3)}_{l,n,U} \underset{l\to\infty}{\to} \widetilde{\beta}^{(3)}_{\infty,n,U}$ by the monotone convergence theorem (note $\left|\widetilde{\beta}^{(3)}_{l,n,U}\right| < \infty$ for all l). Similarly with Alternatives CI1 and CI2, it follows that if $\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}$, then $\widetilde{\beta}^{(3)}_{\infty,n,U}$ cannot be smaller than $\beta_{n,U}$ wpa1 by definition of $\widetilde{\beta}^{(3)}_{\infty,n,U}$ since wpa1, $\alpha_{n,U} \in \Theta \times \widehat{\mathcal{H}}_{\delta}$. Also note $\widetilde{\beta}^{(3)}_{l,n,U} \geq \widetilde{\beta}^{(3)}_{\infty,n,U} - \epsilon$ for arbitrary small number $\epsilon > 0$ for all large l since $\widetilde{\beta}^{(3)}_{l,n,U} \underset{l\to\infty}{\to} \widetilde{\beta}^{(3)}_{\infty,n,U}$. We will let $\kappa_{l,U} = \widetilde{\beta}^{(3)}_{\infty,n,U} - \widetilde{\beta}^{(3)}_{l,n,U}$. Then, we have wpa1, $\widetilde{\beta}^{(3)}_{l,n,U} \geq \beta_{n,U} - \kappa_{l,U}$. Combining these results with Lemma B.5 and noting $\kappa_{l,U} = o_{p,l}(1)$ by construction, similarly with CI1 and CI2, we obtain

$$\begin{aligned}
&P(\beta_0 \leq \widetilde{\beta}^{(3)}_{l,n,U}) + o_{p,l}(1) + o_p(1) \geq P(\beta_{n,U} \leq \widetilde{\beta}^{(3)}_{l,n,U}) + O_p(\kappa_{l,U}) + o_p(1) \\
&\geq P(\beta_{n,U} \leq \widetilde{\beta}^{(3)}_{l,n,U} + \kappa_{l,U}) + o_p(1) \geq P(\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U})
\end{aligned} \tag{55}$$

for all $n \geq \exists N$ and $o_{p,l}(1)$ denotes some random sequences that go to zero as $l \to \infty$.

The remaining proof exactly follows that of Alternative CI1&CI2 and thus this completes the proof of Theorem 5.1 for case (ii) with Alternative CI3. **Case (iii):** It can be proved analogously to case (ii).

Case (iv): Note that analogous results of each (51), (54), and (55) for L in replace of U throughout holds s.t.

$$\begin{aligned}
&P(\widetilde{\beta}^{(1)}_{n,L} \leq \beta_0) \geq P(\widetilde{\beta}^{(1)}_{n,L} \leq \beta_{n,L}) \geq P(\widetilde{c}_{n,L}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,L}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,L}), \\
&P(\widetilde{\beta}^{(2)}_{n,L} \leq \beta_0) + o_p(1) \geq P(\widetilde{\beta}^{(2)}_{n,L} \leq \beta_{n,L}) + o_p(1) \geq P(\widetilde{c}_{n,L}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,L}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,L}), \\
&P(\widetilde{\beta}^{(3)}_{l,n,L} \leq \beta_0) + o_{p,l}(1) + o_p(1) \geq P(\widetilde{\beta}^{(3)}_{l,n,L} \leq \beta_{n,L}) + O_p(\kappa_{l,L}) + o_p(1) \\
&\geq P(\widetilde{c}_{n,L}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,L}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,L})
\end{aligned} \tag{56}$$

by the same argument as in (51), (54), and (55), alternatively. Combining (51), (54), or (55) with (56), we obtain

$$\begin{aligned}
&P(\widetilde{\beta}^{(1)}_{n,L} \leq \beta_0 \leq \widetilde{\beta}^{(1)}_{n,U}) \geq P(\widetilde{\beta}^{(1)}_{n,L} \leq \beta_{n,L}, \beta_{n,U} \leq \widetilde{\beta}^{(1)}_{n,U}) \\
&\geq P(\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}, \widetilde{c}_{n,L}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,L}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,L}), \\
&P(\widetilde{\beta}^{(2)}_{n,L} \leq \beta_0 \leq \widetilde{\beta}^{(2)}_{n,U}) + o_p(1) \geq P(\widetilde{\beta}^{(2)}_{n,L} \leq \beta_{n,L}, \beta_{n,U} \leq \widetilde{\beta}^{(2)}_{n,U}) + o_p(1) \\
&\geq P(\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}, \widetilde{c}_{n,L}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,L}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,L}), \\
&P(\widetilde{\beta}^{(3)}_{l,n,L} \leq \beta_0 \leq \widetilde{\beta}^{(3)}_{l,n,U}) + o_{p,l}(1) + o_p(1) \geq P(\widetilde{\beta}^{(3)}_{l,n,L} \leq \beta_{n,L}, \beta_{n,U} \leq \widetilde{\beta}^{(3)}_{l,n,U}) + O_p(\max\{\kappa_{l,L}, \kappa_{l,U}\}) + o_p(1) \\
&\geq P(\widetilde{c}_{n,U}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,U}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,U}, \widetilde{c}_{n,L}(j,\widehat{\gamma}_{n,j,m},\alpha_{n,L}) \geq 0 \ \forall (j,m) \in \widehat{\mathcal{B}}_{n,L}).
\end{aligned} \tag{57}$$

Analogous results to those of (52)-(53) holds with L in place of U throughout. Note that the $\liminf$ of the RHS of the second inequality in (57) for each alternative CI is at least as large as the RHS of (53) because $\beta_{+,U} = \beta_{+,L}$ implies that $\alpha_{+,U} = \alpha_{+,L}$ and $\mathcal{B}_{+,U} = \mathcal{B}_{+,L}$ by the definitions of $\beta_{+,U}$, $\alpha_{+,U}$, and $\mathcal{B}_{+,U}$ in Assumption B.4 and (46), respectively. We have shown Theorem 5.1 holds for all four cases.

[24]Note that with some abuse of notation, when we consider $\widehat{\mathcal{H}}_{l,\delta}$ as a sequence of sets indexed by l, we treat $\widehat{h}(\cdot,\theta)$ is fixed even though $l \geq n \to \infty$.

B.2.2 Proof of Lemma B.1

For each $j \in \mathcal{I}_J$, let $\mathcal{F}_{\xi_j} = \{\xi_j(y,x,\gamma,\theta,h) = \left(P(y^{(j)}|x,\theta,h) - \mathbf{1}[y=y^{(j)}]\right) q_\gamma(x) - c_0(j,\gamma,\theta,h) : (\gamma,\theta,h) \in \Gamma_{all} \times \Theta \times \mathcal{H}\}$ denote the class of measurable functions indexed by (γ,θ,h). Note that $\widehat{\nu}_n(j,\gamma,\theta,h) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_j(y_i,x_i,\gamma,\theta,h)$. Also note that $E[\xi_j(Y,X,\gamma,\theta,h)] = 0$ for all $(j,\gamma,\theta,h) \in \mathcal{I}_J \times \Gamma_{all} \times \Theta \times \mathcal{H}$ by construction. To prove this lemma, first, we extend Lemma 1 of Chen, Linton, and van Keilegom (2003) to fit our case.

Lemma B.6 *Let $\{Y_i, X_i\}_{i=1}^n$ be iid with $E[\xi_j(Y,X,\gamma,\theta,h)] = 0$ for all $(j,\gamma,\theta,h) \in \mathcal{I}_J \times \Gamma_{all} \times \Theta \times \mathcal{H}_\delta$. Suppose that $\mathcal{F}_{\xi_j} = \{\xi_j(y,x,\gamma,\theta,h) : (\gamma,\theta,h) \in \Gamma_{all} \times \Theta \times \mathcal{H}\}$ is P-Donsker (or satisfies $\int_0^\infty \sqrt{\log N_{[]}\left(\varepsilon, \mathcal{F}_{\xi_j}, \|\cdot\|_{L_2(P)}\right)} d\varepsilon < \infty$); and that $\xi_j(y,x,\gamma,\theta,h)$ is $L_2(P)$-continuous at all $(\gamma,\theta,h) \in \Gamma_{all} \times \Theta \times \mathcal{H}_\delta$. Then, (44) and (50) hold.*

Proof. Noting $\widehat{\nu}_n(j,\gamma,\theta,h) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \xi_j(y_i,x_i,\gamma,\theta,h)$ by definition, (44) is obtained by extending Pakes and Pollard (1989)'s Lemma 2.17 from the case $m(\cdot,\theta)$ to our case $\xi_j(Y,X,\gamma,\theta,h)$. Its proof is essentially the same with theirs. Now (50) is obtained from Giné and Zinn (1990). ■

Now we prove Lemma B.1. From Lemma B.6, it suffices to show $\int_0^\infty \sqrt{\log N_{[]}\left(\varepsilon, \mathcal{F}_{\xi_j}, \|\cdot\|_{L_2(P)}\right)} d\varepsilon < \infty$. Note

$$
\begin{aligned}
&\left|\xi_j(y,x,\gamma_1,\theta_1,h_1) - \xi_j(y,x,\gamma_2,\theta_2,h_2)\right| \\
\le& \left|\xi_j(y,x,\gamma_1,\theta_1,h_1) - \xi_j(y,x,\gamma_1,\theta_2,h_2)\right| + \left|\xi_j(y,x,\gamma_1,\theta_2,h_2) - \xi_j(y,x,\gamma_2,\theta_2,h_2)\right| \\
\le& \left|P(y^{(j)}|x,\theta_1,h_1) - P(y^{(j)}|x,\theta_2,h_2)\right| q_{\gamma_1}(x) + \left|\int \left(P(y^{(j)}|x,\theta_1,h_1) - P(y^{(j)}|x,\theta_2,h_2)\right) q_{\gamma_1}(x) dF_X(x)\right| \\
&+ \left|P(y^{(j)}|x,\theta_2,h_2) - \mathbf{1}[y=y^{(j)}]\right| |q_{\gamma_1}(x) - q_{\gamma_2}(x)|
\end{aligned}
$$

and hence we have

$$
\begin{aligned}
&\left|\xi_j(y,x,\gamma_1,\theta_1,h_1) - \xi_j(y,x,\gamma_2,\theta_2,h_2)\right|^2 \\
\le& 3\left(P(y^{(j)}|x,\theta_1,h_1) - P(y^{(j)}|x,\theta_2,h_2)\right)^2 q_{\gamma_1}(x)^2 + 3\int \left(P(y^{(j)}|x,\theta_1,h_1) - P(y^{(j)}|x,\theta_2,h_2)\right)^2 q_{\gamma_1}(x) dF_X(x) \\
&+ 3\left(P(y^{(j)}|x,\theta_2,h_2) - \mathbf{1}[y=y^{(j)}]\right)^2 (q_{\gamma_1}(x) - q_{\gamma_2}(x))^2 \\
\le& (C_1(X) + C_2)\left(\|\theta_1-\theta_2\|_E^2 + \|h_1-h_2\|_{\mathcal{H}}^2\right) + 3(q_{\gamma_1}(x) - q_{\gamma_2}(x))^2
\end{aligned}
$$

for some $C_1(X) < \infty$ and $C_2 < \infty$ by Assumption 4.4 and the condition (b). Thus, it follows that

$$
\left(E\left[\sup_{\|\gamma_1-\gamma_2\|<\delta, \|\theta_1-\theta_2\|_E<\delta, \|h_1-h_2\|_{\mathcal{H}}<\delta} \left|\xi_j(Y,X,\gamma_1,\theta_1,h_1) - \xi_j(Y,X,\gamma_2,\theta_2,h_2)\right|^2\right]\right)^{1/2} \le C\delta \tag{58}
$$

from the definition of the semi-norm $\|\gamma_1 - \gamma_2\|$. Therefore, $\xi_j(y,x,\gamma,\theta,h)$ is locally uniformly $L_2(P)$-continuous with respect to $(\gamma,\theta,h) \in \Gamma_{all} \times \Theta \times \mathcal{H}_\delta$ by Theorem 6 in Andrews (1994a). The remaining proof is obtained similarly with the proof of the theorem 3 in Chen, Linton, and van Keilegom (2003). Now let $\{\gamma_k : k = 1,\ldots,N_1\}$ be a δ-cover for $(\Gamma_{all}, \|\cdot\|)$, $\{\theta_k : k = 1,\ldots,N_2\}$ be a δ-cover for $\left(\Theta, \|\cdot\|_E\right)$, $\{h_k : k = 1,\ldots,N_3\}$ be a δ-cover for $\left(\mathcal{H}, \|\cdot\|_{\mathcal{H}}\right)$. Also let $\mathcal{N}_1 \equiv \{1,\ldots,N_1\}$, $\mathcal{N}_2 \equiv \{1,\ldots,N_2\}$, and $\mathcal{N}_3 \equiv \{1,\ldots,N_3\}$. Then, by (58), for any $\xi_j(y,x,\gamma,\theta,h)$, there exist $k_1 \in \mathcal{N}_1$, $k_2 \in \mathcal{N}_2$, and $k_3 \in \mathcal{N}_3$ such that $\left|\xi_j(y,x,\gamma,\theta,h) - \xi_j\left(y,x,\gamma_{k_1},\theta_{k_2},h_{k_3}\right)\right|$ is bounded by $\sup_{(\gamma,\theta,h) \text{ s.t. } \|\gamma-\gamma_{k_1}\|<\delta, \|\theta-\theta_{k_2}\|_E<\delta, \|h-h_{k_3}\|_\infty<\delta} \left|\xi_j(y,x,\gamma,\theta,h) - \xi_j\left(y,x,\gamma_{k_1},\theta_{k_2},h_{k_3}\right)\right| \equiv b_j(y,x,\gamma_{k_1},\theta_{k_2},h_{k_3},\delta)$. It follows that $\xi_j\left(y,x,\gamma_{k_1},\theta_{k_2},h_{k_3}\right) - b_j(y,x,\gamma_{k_1},\theta_{k_2},h_{k_3},\delta) \le \xi_j(y,x,\gamma,\theta,h) \le \xi_j\left(y,x,\gamma_{k_1},\theta_{k_2},h_{k_3}\right) + b_j(y,x,\gamma_{k_1},\theta_{k_2},h_{k_3},\delta)$ and that $\left(E\left[b_j(Y,X,\gamma_{k_1},\theta_{k_2},h_{k_3},\delta)^2\right]\right)^{1/2} \le C\delta$ for all $\left(\gamma_{k_1},\theta_{k_2},h_{k_3}\right)$ and all positive sequence tending to zero $\delta = o(1)$. Therefore, an $\varepsilon = 2C\delta$-bracket for $\left(\mathcal{F}_{\xi_j}, \|\cdot\|_{L_2(P)}\right)$ is formed as

$$
\left\{\begin{array}{l} \xi_j\left(y,x,\gamma_{k_1},\theta_{k_2},h_{k_3}\right) - b_j(y,x,\gamma_{k_1},\theta_{k_2},h_{k_3},\delta),\ \xi_j\left(y,x,\gamma_{k_1},\theta_{k_2},h_{k_3}\right) + b_j(y,x,\gamma_{k_1},\theta_{k_2},h_{k_3},\delta) \\ : k_1 \in \mathcal{N}_1,\ k_2 \in \mathcal{N}_2, \text{ and } k_3 \in \mathcal{N}_3 \end{array}\right\}.
$$

It follows that $N_{[]}\left(\varepsilon, \mathcal{F}_{\xi_j}, \|\cdot\|_{L_2(P)}\right) \le N\left(\frac{\varepsilon}{2C}, \Theta, \|\cdot\|_E\right) \times N\left(\frac{\varepsilon}{2C}, \mathcal{H}, \|\cdot\|_{\mathcal{H}}\right) \times N\left(\frac{\varepsilon}{2C}, \Gamma_{all}, \|\cdot\|\right)$. Combining this result, the condition (a), and the arguments following (44), we complete the proof.

B.2.3 Proof of Lemma B.2

We can rewrite

$$
\begin{aligned}
&\sqrt{n}\left(\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})-c_0(j,\gamma_{0,j,m},\widehat{\alpha}_{n,U})\right)\\
&=\sqrt{n}\left(\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})-c_0\left(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U}\right)\right)+\sqrt{n}\left(c_0(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})-c_0(j,\gamma_{0,j,m},\widehat{\alpha}_{n,U})\right)\\
&=\widehat{\nu}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})+\widehat{Z}_n\left(j,m,\widehat{\alpha}_{n,U}\right)
\end{aligned}
$$

using the definitions of $\widehat{v}_n(\cdot,\cdot,\cdot)$ and $\widehat{Z}_n(\cdot,\cdot,\cdot)$ in (41) and (42), respectively. From the results of part (i) and part (ii) in Lemma B.4, we have

$$
\widehat{\nu}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\underset{d}{\to}\nu_0(j,\gamma_{0,j,m},\alpha_{+,U})\text{ and }\widehat{Z}_n\left(j,m,\widehat{\alpha}_{n,U}\right)\underset{d}{\to}Z_0\left(j,m,\alpha_{+,U}\right) \tag{59}
$$

since $\|\widehat{\alpha}_{n,U}-\alpha_{+,U}\|_s\underset{p}{\to}0$ by the condition (i) and Assumption B.4 (ii) and since $\left\|\widehat{\gamma}_{n,j,m}-\gamma_{0,j,m}\right\|\underset{p}{\to}0$ by Assumption 4.7. Finally, note Assumption B.3, Assumption 4.7, and the condition (i) imply that

$$
\begin{aligned}
&\left|\widehat{w}_n\left(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U}\right)-w_0\left(j,\gamma_{0,j,m},\alpha_{+,U}\right)\right|\\
&\le\left|\widehat{w}_n\left(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U}\right)-w_0\left(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U}\right)\right|+\left|w_0\left(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U}\right)-w_0\left(j,\gamma_{0,j,m},\alpha_{+,U}\right)\right|\\
&\le\sup_{(j,\gamma,\alpha)\in\mathcal{I}_J\times\Gamma_{all}\times\mathcal{A}_\delta}\left|\widehat{w}_n(j,\gamma,\alpha)-w_0(j,\gamma,\alpha)\right|+\left|w_0\left(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U}\right)-w_0\left(j,\gamma_{0,j,m},\alpha_{+,U}\right)\right|=o_p(1).
\end{aligned} \tag{60}
$$

Combining (59) and (60), the claim follows.

B.2.4 Proof of Lemma B.3

Consider

$$
\begin{aligned}
&\sqrt{n}\left(\widehat{c}_n^*(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U})-\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right)\\
=\;&\sqrt{n}\left(\widehat{c}_n^*(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U})-\widehat{c}_n(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U})-c_n^*(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})+\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right) &(61)\\
&+\sqrt{n}\left(\widehat{c}_n(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U})-\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})-c_0(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U})+c_0(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right) &(62)\\
&+\sqrt{n}\left(c_n^*(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})-\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right)+\sqrt{n}\left(c_0(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U})-c_0(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right) &(63)
\end{aligned}
$$

Now note that (61) is $o_{P^*}(1)$ by the condition (i) and (iii). Note that from the definition of $\widehat{\nu}_n(\cdot,\cdot,\cdot)$, we have

$$
\begin{aligned}
&\sqrt{n}\left(\widehat{c}_n(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U})-\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})-c_0(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U})+c_0(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right)\\
&=\widehat{\nu}_n(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U})-\widehat{\nu}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U}).
\end{aligned}
$$

and hence (62) is $o(1)$ a.s. by (44) (a.s. version). From Giné and Zinn (1990), we know that

$$
\sqrt{n}\left(c_n^*(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})-\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right)=\sqrt{n}\left(\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})-c_0(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right)+o_{P^*}(1)
$$

and note also that $\sqrt{n}\left(\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})-c_0(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right)=\widehat{\nu}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\underset{d}{\to}\nu_0(j,\gamma_{0,j,m},\alpha_{+,U})$ by the part (i) of Lemma B.4 (note that $\widehat{\alpha}_{n,U}\in\widehat{\mathcal{A}}_n\subset\mathcal{A}_\delta$ wpa1) and since $\|\widehat{\alpha}_{n,U}-\alpha_{+,U}\|_s=o(1)$ *a.s.* by the condition (i) and Assumption B.4 (ii) (a.s. version). From these results, it follows that for the first term in (63), $\sqrt{n}\left(c_n^*(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})-\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right)\underset{d}{\to}\nu_0(j,\gamma_{0,j,m},\alpha_{+,U})+o_{P^*}(1)$. Finally, we note that for the second term in (63), $\sqrt{n}\left(c_0(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U})-c_0(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right)\underset{d}{\to}Z_0\left(j,m,\alpha_{+,U}\right)+o_{P^*}(1)$ by the part (ii) of Lemma B.4 and since $\|\widehat{\alpha}_{n,U}-\alpha_{+,U}\|_s=o(1)$ *a.s.*. Therefore, we have

$$
\sqrt{n}\left(\widehat{c}_n^*(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U})-\widehat{c}_n(j,\widehat{\gamma}_{n,j,m},\widehat{\alpha}_{n,U})\right)\underset{d}{\to}\nu_0(j,\gamma_{0,j,m},\alpha_{+,U})+Z_0\left(j,m,\alpha_{+,U}\right)+o_{P^*}(1). \tag{64}
$$

Now it remains to show that $\widehat{w}_n^*\left(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U}\right)=w_0\left(j,\gamma_{0,j,m},\alpha_{+,U}\right)+o_{P^*}(1)$. This holds because

$$
\begin{aligned}
&\left|\widehat{w}_n^*\left(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U}\right)-w_0\left(j,\gamma_{0,j,m},\alpha_{+,U}\right)\right|\\
&\le\left|\widehat{w}_n^*\left(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U}\right)-\widehat{w}_n\left(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U}\right)\right|+\left|\widehat{w}_n\left(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U}\right)-w_0\left(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U}\right)\right|\\
&\quad+\left|w_0\left(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U}\right)-w_0\left(j,\gamma_{0,j,m},\alpha_{+,U}\right)\right|\\
&\le\sup\nolimits_{(j,\gamma,\alpha)\in\mathcal{I}_J\times\Gamma_{all}\times\mathcal{A}_\delta}\left|\widehat{w}_n^*\left(j,\gamma,\alpha\right)-\widehat{w}_n\left(j,\gamma,\alpha\right)\right|+\sup\nolimits_{(j,\gamma,\alpha)\in\mathcal{I}_J\times\Gamma_{all}\times\mathcal{A}_\delta}\left|\widehat{w}_n\left(j,\gamma,\alpha\right)-w_0\left(j,\gamma,\alpha\right)\right|\\
&\quad+\left|w_0\left(j,\widehat{\gamma}_{n,j,m}^*,\widehat{\alpha}_{n,U}\right)-w_0\left(j,\gamma_{0,j,m},\alpha_{+,U}\right)\right|=o_P^*(1)
\end{aligned}
$$

where the last result holds by Assumptions B.3 and 4.7 (a.s. version), by the condition (ii), and by $\|\widehat{\alpha}_{n,U}-\alpha_{+,U}\|_s=o(1)$ *a.s.*. Therefore, from this result and (64), the claim follows.

References

Ai, C. and X. Chen (2003), "Efficient Estimation of Models With Conditional Moment Restrictions Containing Unknown Functions," *Econometrica*, v71, No.6, 1795-1843.

Aguirregabiria and Mira (2003), "Sequential Simulated-Based Estimation of Dynamic Discrete Games," Boston University Working Paper.

Andrews, D.W.K. (1994a), "Empirical Process Methods in Econometrics," *Handbook of Econometrics* IV, edited by R.F. Engle and D.L. McFadden, 2247-2294.

Andrews, D.W.K. (1994b), "Asymptotics for Semi-parametric Econometric Models via Stochastic Equicontinuity", *Econometrica*, 62, 43-72.

Andrews, D.W.K., S. Berry, and P. Jia (2004), "Confidence Regions for Parameters in Discrete Games with Multiple Equilibria, with an Application to Discount Chain Store Location," Working paper, Yale University.

Aradillas-Lopez, A. (2005), "Semiparametric Estimation of a Simultaneous Game with Incomplete Information," Job Market Paper, Berkeley.

Bajari, P., C.L. Benkard, and J. Levin (2004), "Estimating Dynamic Models of Imperfect Competition," Working Paper, Department of Economics, Duke University.

Berry, S., M. Ovstrovsky, and A. Pakes (2003), "Simple Estimators for the Parameters of Discrete Dynamic Games (With Entry/Exit Examples)," Harvard University Working Paper.

Bjorn, P. and Q. Vuong (1984), "Simultaneous Models for Dummy Endogenous Variables: A Game Theoretic Formulation with an Application to Household Labor Force Participation," Working paper, California Institute of Technology.

Bjorn, P. and Q. Vuong (1985), "Econometric Modelling of a Stackelberg Game With an Application to Household Labor Force Participation," Working paper, California Institute of Technology.

Bresnahan, T.F. and P.C. Reiss (1990), "Entry in Monopoly Markets," *Review of Economic Studies*, 57, p.531-553.

Brenahan, T.F. and P.C. Reiss (1991), "Empirical Models of Discrete Games," *Journal of Econometrics*, 48, p.57-81.

Brock, W. and S. Durlauf (2001), "Interactions-based Models," in: J. Heckman and E. Leamer, eds., *Handbook of Econometrics*, V.5, 3297-3380.

Brock, W. and S. Durlauf (2003), "Multinomial Choice with Social Interactions," Working paper, University of Wisconsin.

Chen, X. (2005), "Large Sample Sieve Estimation of Semi-Nonparametric Models," forthcoming in *Handbook of Econometrics*, V.6.

Chen, X., O. Linton, and I. van Keilegom (2003), "Estimation of Semiparametric Models When the Criterion Function is Not Smooth," *Econometrica*, 71, 1591-1608.

Ciliberto, F. and E. Tamer (2003), "Market Structure and Multiple Equilibrium in Airline Markets," Working paper, Princeton University.

Giné, E. and J. Zinn (1990), "Bootstrapping General Empirical Measures," *Annals of Probability*, 18, 851-869.

Horowitz, J. and Manski, C. (1995), "Identification and Robustness with Contaminated and Corrupted data," *Econometrica*, 63, 281-302.

Imbens, G. and C. Manski (2004), "Confidence Intervals for Partially Identified Parameters," *Econometrica*, 72, 1845-1857.

Kim, K. (2006), "Semiparametric Estimation of Signaling Games", Ph.D. Dissertation, UCLA.

Kooreman, P. (1994), "Estimation of Econometric Models of Some Discrete Games," *Journal of Applied Econometrics*, 9, 255-268.

Manski, C.F. (1990), "Nonparametric Bounds on Treatment Effects," *American Economic Review*, 80, 319-323.

Manski, C.F. and E. Tamer (2002), "Inference on Regressions with Interval Data on a Regressor or Outcome," *Econometrica*, 70, 519-546.

Newey, W.K. and D. McFadden (1994), "Large Sample Estimation and Hypothesis Testing," *Handbook of Econometrics* IV, edited by R.F. Engle and D.L. McFadden, 2247-2294.

Pesendorfer, M. and P. Schmidt-Dengler (2003), "Identification and Estimation of Dynamic Games," NBER working paper 9726.

Pakes, A. and D. Pollard (1989), "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027-1057.

Pollard, D. (1984), *Convergence of Stochastic Processes*. New York: Springer-Verlag.

Pollard, D. (1990), *Empirical Processes: Theory and Applications*, CBMS Conference Series in Probability and Statistics, Vol.2. Hayward, CA.

Seim, K. (2002), "An Empirical Model of Firm Entry with Endogenous Product-type Choices," mimeo, Graduate School of Business, Stanford University.

Shen, X. (1997), "On Methods of Sieves and Penalization," *Annals of Statistics*, 25, 2555-2591.

Shen, X. (1998), "On the Method of Penalization," *Statistica Sinica*, 8, 337-357

Sutton, J. (2000), *Marshall's Tendencies: What Can Economists Know?*, MIT Press, Cambridge adn London.

Sweeting, A. (2003), "Coordination Games, Multiple Equilibria, and the Timing of Radio Commercials," Ph.D. Job Market Paper, MIT.

Tamer, E.T. (2003), "Incomplete Simultaneous Discrete Response Model with Multiple Equilibria," *Review of Economic Studies* 70, p.147-165.

Van der Vaart, A. and J. Wellner (1996),*Weak Convergence and Empirical Processes: with Applications to Statistics*, New York: Springer-Verlag.